

Smyth, Edward (2000) *Raman optical activity of proteins and glycoproteins*. PhD thesis.

<http://theses.gla.ac.uk/5920/>

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Raman Optical Activity of Proteins and Glycoproteins

By

Edward Smyth

A Thesis presented in partial fulfilment for
the degree of Doctor of Philosophy in the
Faculty of Science of the University of Glasgow

Chemistry Department

March 2000

© Edward Smyth

Acknowledgements

I would like to take this opportunity to thank my supervisor Professor Laurence D. Barron for the help and encouragement he gave me during the course of this project.

I would also like to thank my co-supervisor Dr. Lutz Hecht, together with Dr. Alasdair F. Bell, Dr. Ewan W. Blanch and Mr. Christopher D. Syme for their support and occasional bad jokes.

Finally, I thank the Engineering and Physical Sciences Research Council for the award of a Research Studentship.

Abbreviations

b	Bend
BSA	Bovine serum albumin
CCD	Charge-coupled device
CD	Circular dichroism
CID	Circular intensity difference
ConA	Concanavalin A
DCP	Dual circular polarisation
DMSO	Dimethyl sulphoxide
EOM	Electro-optic modulator
FTIR	Fourier transform infrared
GUROAS	Glasgow University Raman optical activity spectrometer
HSNPF	Holographic super notch-plus filter
HSA	Human serum albumin
IgG	Immunoglobulin G
ICP	Incident circular polarisation
IR	Infrared
ip	In-plane
NMA	N-methylacetamide
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect (enhancement)
OR	Optical rotation
ORD	Optical rotatory dispersion
op	Out-of-plane bend
PDB	Protein databank
PPII	Polyproline II
ROA	Raman optical activity
RNA	Ribonucleic acid
SCP	Scattered circular polarisation
SNR	Signal to noise ratio
SDFF	Spectroscopically determined force field

SMF	Stokes Muller formalism
s	Stretch
t	Torsion
TIM	Triose phosphate isomerase
UV	Ultraviolet
UVCD	Ultraviolet circular dichroism
VCD	Vibrational circular dichroism
VOA	Vibrational optical activity

Table of Contents

Abstract

1. Introduction

1.1. Electronic Optical Activity	2
1.2. Vibrational Optical Activity (VOA)	5
1.2.1. Vibrational Circular Dichroism (VCD)	5
1.2.2. Raman Optical Activity (ROA)	7
1.2.3. The Two-group Model	9
1.2.4. Analysis of ROA and VCD	11
1.3. Discussion	14

2. Theory of ROA

2.1. Molecular Property Tensors	16
2.2. ROA Observables	19
2.3. Natural Raman Optical Activity	21
2.4. Backscattering ROA	25
2.5. Ab-initio ROA Calculations	27

3. Instrumentation

3.1. Optical Layout	29
3.2. Detection System (CCD)	34
3.3. Computer Control	36
3.4. Artefact Control	37
3.5. Miscellaneous	39
3.6. Sample Sources and Experimental Conditions	42

4. The Structure of Proteins and Glycoproteins

4.1. Protein Structure Hierarchy	43
4.1.1. Glycoproteins	48
4.2. Normal Modes of Proteins	52

5. Immunoglobulin G and Other Proteins with Mostly β-sheet Structure	
5.1. Immunoglobulin G (IgG)	56
5.2. Avidin	65
5.3. Pepsin	69
5.4. Discussion	73
6. Serine Proteinase Inhibitors	
6.1. Serine Proteinase Inhibitors (serpins)	75
6.1.1. The Active Loop	77
6.1.2. Some Medical Conditions Associated with Dysfunctional Serpins	79
6.2. Chicken and Turkey Ovomucoid	81
6.1. Ovalbumin and Recombinant α_1 -Antiproteinase	86
6.4. Discussion	90
7. Proteins with Irregular Folds	
7.1. Unordered Poly(L-lysine) and Poly(L-glutamic acid)	92
7.2. Phosvitin	95
7.3. Invertase	97
7.4. Bowman-Birk Proteinase Inhibitor	99
7.5. Metallothionein	102
7.6. Discussion	105
8. Some ROA Spectra of Proteins and Carbohydrates	
8.1. Some Protein ROA Spectra	107
8.1.1. Ribonuclease A and B	107
8.1.2. Insulin	108
8.1.3. Ovotransferrin (Conalbumin)	114
8.1.4. Bovine serum albumin (BSA), Human serum albumin (HSA) and glycosylated (?) α_1 -Antiproteinase	119
8.1.5. Aldolase	123
8.1.6. Concanavalin A	127

8.1.7. Type I Dehydroquinase	130
8.2. Ordered and Disordered Polysaccharides	131
8.2.1. Polysaccharides	131
8.2.2. Dextran, Glycogen, Pullulan and Laminarin	132
8.3. Discussion	134
References	137

Abstract

Raman optical activity (ROA), measured in this project as a small difference in the intensity of Raman scattering from chiral molecules in right- and left-circularly polarised incident laser light, offers the potential to provide more information about the structure of biological molecules in aqueous solution than conventional spectroscopic techniques.

Chapter one contains a general discussion of the relative merits of different spectroscopic techniques for structure determination of biomolecules, as well as a brief introduction to ROA. In Chapter two a theoretical analysis of ROA is developed, which extends the discussion in chapter one.

The spectrometer setup and sample preparation is then discussed in chapter three. Instrument and sample conditions are monitored to ensure that the best results are obtained. As with any experimental project problems occur, which may result in a degradation of the spectra obtained. The cause of these problems was explored and remedied whenever possible.

Chapter four introduces a brief account of protein, glycoprotein and carbohydrate structure and function, with a particular emphasis on the structure of proteins.

In the remaining chapters experimental ROA results on proteins and glycoproteins, with some carbohydrate samples, from a wide range of sources are examined. For example, in chapter five some β -sheet proteins are examined. Structural features in these proteins are examined in the extended amide III region of their ROA spectra, revealing that ROA is sensitive to the rigidity or flexibility inherent in proteins.

Chapter six concentrates on a group of proteins (usually glycoproteins) known as the serine proteinase inhibitors (serpins). Medically, the serpins are one of the most important groups of proteins of current interest, with wide-ranging implications in conditions such as Down's syndrome, Alzheimer's disease, and emphysema with associated cirrhosis of the liver. With favourable samples and conditions ROA may offer the potential to understand the mode of function of these proteins in the solution state.

Proteins with irregular folds are the subjects of study in chapter seven. Model homopolypeptides are examined and used to help determine that certain proteins are composed of largely disordered regions of their polypeptide backbone structure in solution. Results from ROA suggest that there are two distinct types of disorder in these proteins. For example, phosvitin appears to possess ‘unstructured’ disorder similar to that of reduced or molten globule proteins; whereas other samples such as invertase and the Bowman-Birk proteinase inhibitor possess ‘structured’ disorder similar to that found in loops and turns.

In chapter eight ROA studies on a set of miscellaneous proteins together with a few carbohydrates are presented. Samples such as aldolase and concanavalin A are potentially useful in revealing the differences found in proteins with differing β -barrel structures. In future studies such comparisons may prove useful in a more general setting of solution structure studies of proteins that are not amenable to conventional structure elucidation techniques such as NMR (due to size constraints) and X-ray crystallography (due to difficult sample preparation). Also, some carbohydrate samples are examined to explore the conditions needed to determine the glycan structure of glycoproteins.

Overall, past and present studies suggest that ROA offers an additional tool for the structure elucidation of biological samples that are difficult to study using standard techniques. Improvements in instrumentation and sample conditions together with numerical and analytical methods such as pattern recognition or regression analysis may provide substantial results in future studies.

Chapter 1

Introduction

The determination of solution structure and dynamics of biomolecules is at the forefront of biomolecular science (Evans, 1995). Thousands of new biopolymers such as proteins, carbohydrates and nucleic acids are discovered yearly, with most of these molecules remaining structurally uncharacterised. Consequently, the need for simple, easily interpretable spectroscopic techniques (vibrational or otherwise) in biochemistry is now more important than ever.

In this respect the central techniques of structural biology, namely X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, are relatively slow methods and are subject to severe limiting conditions on the samples being investigated (Evans, 1995; Lesk, 1991). For example the need to produce “good” crystals in X-ray crystallography that diffract to high resolution is an acquired art and is far from routine (Lesk, 1991). Also, diffraction data quality can suffer from the effects of large B factors (temperature factors), which often results in protein surface loops being excluded from refined model structures (Lesk, 1991). Glycoproteins are particularly difficult to study since glycosylation exacerbates the problems of crystal production and data interpretation (David, 1997).

In some ways NMR can be used to circumvent these problems and gives complementary information to X-ray data (Evans, 1995). Solution state structure characterisation of proteins is now routinely done. However, a few problems remain with the technique such as the upper size limit of biological molecules studied. In NMR experiments nuclear Overhauser enhancements (NOEs) restrict the size of molecules being studied. As the molecular weight of a biomolecule increases above ~25,000 Dalton the NMR signals from chemically distinct groups begin to overlap and become difficult to resolve, resulting in ambiguous structure assignments (Evans, 1995). Also, as the protein models from NMR are based on distance constraints, unique polypeptide structures are not obtained. Modelling refinement procedures can fit NOE data to an ensemble of protein architectures. Despite these disadvantages, X-ray and NMR methods have nonetheless been of immense value because they provide structures of biomolecules at atomic resolution (Evans, 1995; Lesk, 1991).

In this project the novel chiroptical spectroscopic technique of Raman optical activity (ROA) (Barron, 1982) has been used to study proteins that are considered “difficult” samples for X-ray crystallography and NMR. In particular, large glycoproteins and proteins such as immunoglobulin G (IgG) are studied in their native states. With X-ray crystallography the separate fragments of this molecule are routinely studied (there are quite a few fragments depending on how the molecule is treated enzymatically) with only a few intact IgG molecules providing structural information (Marquart & Deisenhofer, 1982). All of this raises the question: how do native proteins differ from their chemically modified variants, and if so, can these differences be seen by spectroscopic techniques such as ROA. In the sections that follow ROA will be compared with the conventional chiroptical techniques of optical rotation and circular dichroism. “Chiroptical” refers to the use of circularly polarised radiation in order to be sensitive to molecular chirality.

1.1. Electronic Optical Activity

The differential refraction and absorption of left- and right-circularly polarised light through a medium are referred to as optical rotation (OR) and circular dichroism (CD), respectively (Barron, 1982). For samples of chiral molecules in the absence of static external fields, using visible or ultraviolet light, these effects are manifestations of natural electronic optical activity (Vosolov & Woody, 1994).

If an optically active medium has different refractive indices for left- and right-circularly polarised light, the light will travel at different speeds through the medium with respect to different polarisation states (Vosolov & Woody, 1994). As illustrated in Figure 1.1, this will cause the plane of plane-polarised light to rotate through an angle α (Barron, 1982; Vosolov & Woody, 1994).

In a similar manner, circular dichroism (CD) occurs if the wavelength in the visible or UV region of the electromagnetic spectrum coincides with the absorption band of a chromophore in a chiral molecule (Snatzke, 1994). The variation of OR through an absorption band, called optical rotatory dispersion (ORD) was widely used for stereochemical studies in the 1950s (Moffitt & Yang, 1956).

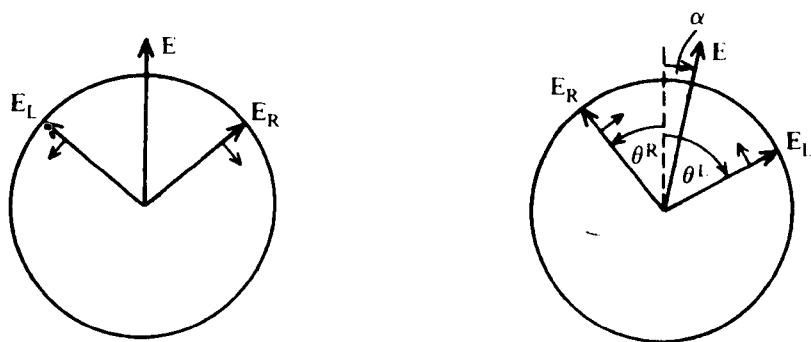


Figure 1.1 Optical rotation of resultant electric field vector E . E_R and E_L are the right and left electric field components of circularly polarised light. (Reproduced from Barron, 1982.)

However, due to advances in instrumentation over the past twenty to thirty years CD has replaced ORD and has developed into the central chiroptical technique of protein structure elucidation (Rogers & Nordén, 1997). UV absorption and associated chiral effects originate from the displacement of electronic charge in a chromophore (Rogers & Nordén, 1997). If this movement of charge is constrained to a helical path, right- and left-circularly polarised light will interact differently with a right- or left-helical electronic current. For optical activity (OA) to exist in a molecule there must be no centres or planes of symmetry, i.e. no improper rotation axes (Chang, 1971). With this in mind the chromophores in CD studies can be either intrinsically asymmetric (e.g. hexahelicenes) or can have induced asymmetry from perturbations in the local molecular environment. For example, in 3-methylcyclohexanone the methyl group induces asymmetry in the carbonyl group (the local unperturbed symmetry in the $>C=O$ group is C_{2v} implying the existence of a reflection plane) of this molecule. As a general rule this induced asymmetry is very weak (Chang, 1971). The main carbonyl CD band involves a $\pi^* \leftarrow n$ transition, which is magnetic dipole allowed (along the z -axis of the $C=O$ bond) but electric dipole forbidden on symmetry grounds (Lightner, 1994). However, the electric dipole forbidden transition is induced by the chiral environment of a molecule and is observed in a potentially troublesome region of a CD spectrum as bands due to amino acid side chains are also observed here (Rogers & Nordén, 1997; Woody, 1994), which can confuse attempts to extract the protein secondary structure content percentage. This is especially true of proteins with short stretches of α -helix \sim four residues in length, resulting in 'end-effects' (Rogers & Nordén, 1997; Woody, 1994). The rotatory strength, R , can be used to give a measure

of CD band intensities (Barron, 1982; Vosolov & Woody, 1994; Chang, 1971). For the $j \leftarrow n$ transition

$$R_{jn} = \text{Im} \langle n | \mu_\alpha | j \rangle \langle j | m_\alpha | n \rangle \quad (1.1.1)$$

In protein CD experiments the majority of information comes from amide $\pi^* \leftarrow n$ and $\pi^* \leftarrow \pi$ transitions in the far UV range ~ 220 to 190 nm, which yields information about α -helix, β -sheet and random coil (Rogers & Nordén, 1997; Woody, 1994). Information on the protein tertiary structure is deduced from aromatic $\pi^* \leftarrow \pi$ transitions in the near UV ~ 250 to 290 nm (Rogers & Nordén, 1997; Woody, 1994). Here bathochromic (red shift) effects (Williams & Fleming, 1989) are observed when aromatic residues move from a hydrophobic to a hydrophilic environment, that is, on going from a native tertiary fold to a denatured state. This is explained by using the Frank-Condon principle (Williams & Fleming, 1989; Atkins, 1997), which states that during an electronic transition atoms do not move; however, electrons, including solvent molecule electrons, may reorganise. This will often result in excited states more polar than the ground state, resulting in dipole-dipole interactions that lower the energy of the excited state more than the ground state, with small red shifts of ~ 10 to 20 nm being observed (Woody, 1994). In addition, aromatic CD bands from tryptophan may vanish, due to the relative abundance of this residue in proteins, prohibiting this avenue of analysis (Rogers & Nordén, 1997; Woody, 1994).

A disadvantage of CD is its lack of spectral resolution: most CD bands are broad and much spectral information is lost in overlapping transitions (Diem, 1993). Consequently, deconvolution methods are used to untangle complex spectra (Diem, 1993), which can introduce “electronic distortions”. A good example of this problem is the determination of β -turn (Rogers & Nordén, 1997; Woody, 1994). Although “disordered” polypeptides give a clear CD band, individual β -turn types (sometimes classed as disordered) are not recognised. Another example is the discrimination between α - and 3_{10} -helix (3_{10} -helix is an alternative description of a type III β turn (Woody, 1994), which itself is a variant on type I β -turn). The situation here is complicated by the fact that, although model CD calculations exist for 3_{10} -helices in proteins (Woody, 1994), experimental CD data of ‘isolated’ model 3_{10} -helices exist only for Aib types (Toniolo *et al.*, 1996). Typically, both classes of helix are classed

together spectroscopically (Woody, 1994).

1.2. Vibrational Optical Activity (VOA)

The natural electronic optical activity discussed in the previous section can be extended to the vibrational spectrum by introducing two new forms of chiroptical spectroscopy known as Raman optical activity (ROA) (Barron, 1982) and vibrational circular dichroism (VCD) (Diem, 1993).

Both techniques were developed in the late 1960s and early 1970s with ROA first predicted by Atkins and Barron in (1969) and later given a more definitive treatment by Barron and Buckingham in (1971). The experimental confirmation of ROA was provided by Barron *et al.* in (1973) with VCD following soon after by Holzwarth *et al.* (1974) and Nafie *et al.* (1975).

The main benefit of using VOA over UVCD stems from the $3N - 6$ normal vibrational modes that are potentially accessible in the vibrational spectrum of a biological molecule (Barron & Hecht, 1994). In a CD experiment stereochemical information can only be obtained if a chromophore is present (Rogers & Nordén, 1997; Barron & Hecht, 1994). Of course not all vibrational modes are accessible in a given experimental setup (Polavarapu, 1998). But even with this restriction a typical vibrational experiment will have many active modes from different parts of a molecule, and information on many parts of a biopolymer will be contained within the data supplied in a vibrational spectrum (Barron & Hecht, 1994; Polavarapu, 1998). As well as this advantage both VOA experiments are similar to other chiroptical methods in their sensitivity to the chiral framework of a molecule:- both will give information about the vibrational modes that are directly involved with chiral elements (Barron & Hecht, 1994; Polavarapu, 1998).

1.2.1. Vibrational Circular Dichroism (VCD)

VCD spectra are obtained directly by measuring the differential absorption of left- and right-circularly polarised infrared (IR) radiation to give, after normalisation,

the dissymmetry factor (Barron, 1982)

$$g = \frac{(\epsilon^L - \epsilon^R)}{\frac{1}{2}(\epsilon^L + \epsilon^R)} \quad (1.2.1)$$

with ϵ^L and ϵ^R the extinction coefficients in left- and right-circularly polarised IR radiation.

Within a few years of its discovery VCD was in routine use as a technique for structure elucidation, due in large part to the availability of suitable spectroscopic instrumentation (Diem, 1993). However, the interpretation of early VCD results were relatively unsuccessful owing to a lack of a generally applicable theory of the phenomenon (Diem, 1993) with most studies at the time depending on empirical correlations. Recently the situation has improved with theories being developed for small molecules (Nafie & Freeman, 1983; Devlin & Stephens, 1999); however, the situation for proteins is still empirically based. Empirical correlations can be made with regression analysis techniques being applied to data from the accessible regions of a VCD spectrum (Keiderling, 1996). In the mid to late 1980s some of the first polypeptide and protein spectra were produced by Keiderling's laboratory (Diem, 1993), namely those of poly- γ -benzyl-L-glutamate, a right-handed α -helical molecule in CHCl_3 ; and, myoglobin, hemoglobin, papain, ribonuclease A, lysozyme, and chymotrypsin (Diem, 1993). These studies have shown that VCD and CD are in close agreement with each other, giving similar information on the amount of protein secondary structure (Diem, 1993; Keiderling, 1996).

In some ways this is disappointing, as it would suggest that VCD offers little advantage over CD. Also, VCD is currently limited to the 6 μm spectral region (Diem, 1993; Polavarapu & Deng, 1996), corresponding to the Amide I and Amide II regions and, to have any real advantage over CD, greater spectral coverage is required. Recently, this problem has been partly remedied with lower wavenumber regions made more accessible (Polavarapu & Deng, 1996). The Amide III region of a VCD spectrum can be measured but is normally sacrificed as the amide I' is strongly favoured (Keiderling, 1996). Presently, VCD measurements are made almost exclusively in D_2O (Keiderling, 1996) which has the disadvantage of producing spectra of proteins with differing amounts of deuteration in different regions of the

polypeptide structure, depending on the accessibility of exchangeable protons to the solvent (Evans, 1995). Consequently, the interpretation of a protein spectrum is made more difficult. This problem arises from the fact that water, the natural medium of a protein, is a strong IR absorber (Keiderling, 1996).

1.2.2. Raman Optical Activity (ROA)

ROA differs from the one photon absorption process of VCD by being a two photon inelastic scattering process (Diem, 1993). ROA is defined as the difference in the intensity of Raman scattered light from chiral molecules in right- and left-circularly polarised incident light (Barron, 1982; Barron & Hecht, 1994). A convenient experimental quantity is the dimensionless circular intensity difference (Barron, 1982; Barron & Hecht, 1994)

$$\Delta = \frac{(I^R - I^L)}{(I^R + I^L)} \quad (1.3.1)$$

where I^R and I^L are the scattered intensities in right- and left-circularly polarised incident light, respectively. The numerator is the measured ROA signal and the denominator the parent Raman signal, which acts as a normalisation condition to standardise the ROA observable and make it compatible with measurements on other spectrometers.

A brief digression is now needed to explain the vibrational analysis of proteins, from Raman spectra. Here, two main classes of signals are observed: side chain vibrations from amino acid residues (Lord & Yu, 1970) and amide linkage vibrations from the protein backbone. Aromatic residues such as tryptophan, phenylalanine and tyrosine together with the sulphur containing residues methionine and cystine make up a large part of a typical protein Raman spectrum (Lord & Yu, 1970; Painter & Koenig, 1975; Pézolet *et al.*, 1976). Also, aliphatic CH₂ and CH₃ deformations are easily observed and of importance if quantitative information on the β -sheet secondary structure is required (Pézolet *et al.*, 1976). Here the CH₂ deformation Raman band at $\sim 1450 \text{ cm}^{-1}$ is used as an internal intensity standard in the

method used by Pézolet *et al.* (1976) and independently by Lippert *et al.* (1976). The method is only applicable if a well-defined band at $\sim 1240\text{ cm}^{-1}$ is observed, which corresponds to the β -sheet region of a Raman spectrum. Here, X-ray data on the β -sheet contents of a number of reference proteins is used to plot a calibration curve of β -sheet content against the relative intensity of the 1240 cm^{-1} Raman band to the 1450 cm^{-1} CH_2 Raman band, times the average number of CH_2 groups per residue (Pézolet *et al.*, 1976). Finally, the hydrogen bonding of the aromatic residues tyrosine and tryptophan can be followed by observing changes in the intensities of bands assigned to these residues. In particular, the Fermi resonance of tyrosine (arising from a ring-breathing mode and an overtone of an out-of-plane bending vibration) at ~ 830 to 850 cm^{-1} is of interest (Tu, 1986; Carey, 1982). The ratio, R_{Tyr} , of the 850 cm^{-1} line to the 830 cm^{-1} line acts as an indicator of “buried” versus “exposed” tyrosines, which is sensitive to the hydrogen bonding or state of ionisation of the phenolic hydroxyl group (Carey, 1982). A ratio near 0.3 indicates that the OH is strongly bound to a negative acceptor; a value of 1.25 suggests moderately strong hydrogen bonding to H_2O and 2.5 if the OH is a strong hydrogen bond acceptor.

The main regions of interest in an ROA and a Raman spectrum are given in table 1.1 with only the backbone and amide vibrations indicated (Diem, 1993; Krimm & Bandekar, 1986).

ROA directly probes the polypeptide backbone conformation, as it is sensitive to the chiral conformation, as well as some side chain conformations (Barron & Hecht, 1994; Barron *et al.*, in press); whereas, Raman spectroscopy gives a superposition of bands from the side chains (Lord & Yu, 1970) and the polypeptide backbone.

This difference in sensitivity to the polypeptide backbone secondary structure is demonstrated by the work of Lord and Yu (1970). In the late 1960s and early 1970s these researchers pioneered some of the earliest work on protein Raman spectroscopy, with initial interest on lysozyme. In Figure 1.2, Raman spectra of lysozyme (top spectrum) and of its constituent amino acids (bottom spectrum) are presented for comparison (Lord & Yu, 1970).

Table 1.1 Vibrational modes of the amide linkage.

Amide mode	Frequency (cm ⁻¹)	Vibrational components
Amide I	1630 to 1700	C=O _s , C _α -N _s and N-H _{ip}
Amide II	~1560	N-H _{ip} and C _α -N _s
Amide III (extended)	1230 to 1350	N-H _{ip} and C-H _b
Amide IV	Various	C=O _{ip}
Amide V	~725	C _α -N _t and N-H _{op}
Amide VI	Various	C=O _{op}
Skeletal stretch	880 to 960	Backbone C _α -C _s and C _α -N _s

b=bend, ip=in-plane, op=out-of-plane bend, s=stretch, t=torsion. (Reproduced from Diem, 1993.)

The spectra of lysozyme (a) and its constituent amino acids (b) were obtained at pHs ~ 5.2 and ~ 1, respectively (Lord & Yu, 1970), and are both dominated by aromatic bands. A glance at these spectra reveals that the lysozyme Raman spectrum is similar to the Raman spectrum of its constituent amino acids, demonstrating that side chain vibrations dominate the Raman spectra of proteins. This is in contrast to the differences that are seen in comparing a ROA spectrum, of a protein, to its parent Raman spectrum (see examples in other chapters). Chiral secondary structure elements such as α-helices and β-sheets are probed directly by ROA, with side chain vibrations making less dominant contributions to an ROA spectrum when compared with its parent Raman spectrum (Barron & Hecht, 1994).

1.2.3. The Two-Group Model

The generation of ROA is easily visualised by using a two-group model (Barron, 1982; Polavarapu, 1998; Barron & Buckingham, 1974; Stone, 1976). In this model two achiral axially symmetric groups are held in a chiral arrangement with respect to each other (Barron, 1982; Barron & Buckingham, 1974). As illustrated in Figure 1.3, ROA is generated through the interference of photons scattered independently from the two groups.

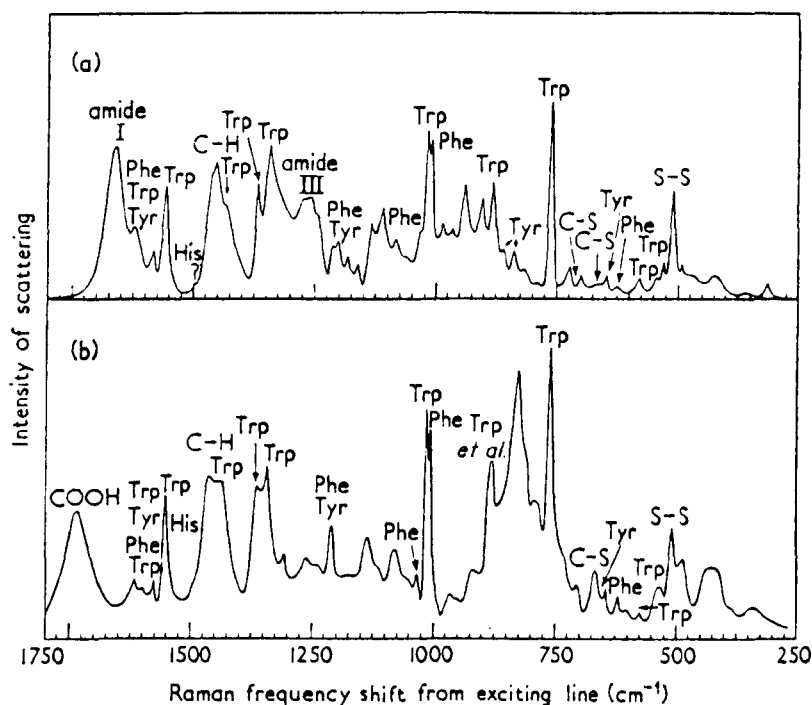


Figure 1.2 Raman spectra of lysozyme (a) and its constituent amino acids (b). (Reproduced from Lord & Yu, 1970.)

When compared with optical rotation, ROA signal generation is much easier to visualise. In OR the Kirkwood model of dynamic scattering is invoked, and requires the interference of an unscattered transmitted photon (in the forward direction) with a scattered photon that has sampled the overall chirality of the structure by being deflected from one group to another (Kirkwood, 1937). In ROA the unscattered photons are unimportant (Barron & Buckingham, 1974).

The two-group model provides an explanation for the poor performance of VCD in the Amide I region when compared with ROA (Barron, 1982). It can be shown that the idealised twisting modes and deformation modes localised on each group are generally more favourable for ROA (Barron, 1982). Here symmetric and antisymmetric combinations of internal coordinates corresponding to the deformation of the angles φ_1 and φ_2 between the group axes and the connecting bond give at some instant the general bond angle $\varphi_1 + \Delta\varphi_1$ and $\varphi_2 + \Delta\varphi_2$ for two vibrations. Note that for this idealised case $\varphi_1 = \varphi_2 = 90^\circ$. Deformation modes require a change in the intrinsic dipole moment μ_i with respect to a change in the deformation angle $\Delta\varphi$, yielding $(\partial\mu_i/\partial\Delta\varphi)_0 \neq 0$, to be VCD active (Barron, 1982). However, in a first approximation, for CH deformations, there is no change in the intrinsic dipole

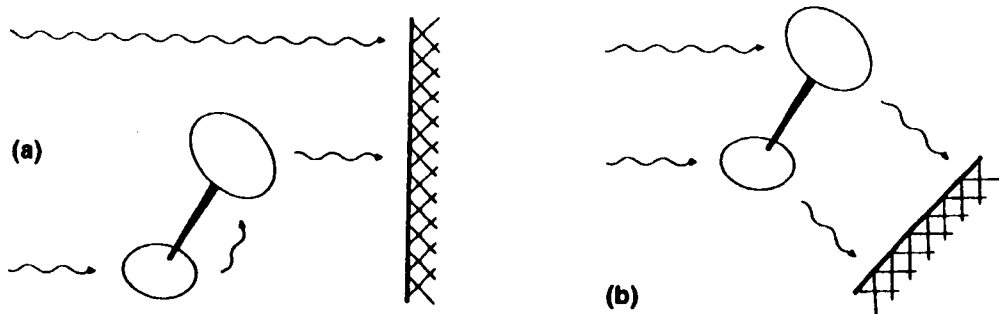


Figure 1.3 Two-group model of (a) optical rotation and (b) Raman optical activity. (Reproduced from Barron & Hecht, 1994.)

moment μ_i with $\Delta\varphi$ (Barron, 1982). On the other hand, at the same level of approximation, the ROA is non-zero, i.e. even when $(\partial\alpha_i / \partial\Delta\varphi)_0 = 0$. The following section explains these results in more detail. In fact, ROA is generally more favourable than VCD since ROA Δ values are larger than VCD g values by $\omega / \omega_p (= \lambda_p / \lambda$ since $\omega = 2\pi c / \lambda$) where ω is the exciting frequency and ω_p is the vibrational frequency (Barron, 1982). Selecting values for $\lambda_p = 50000$ nm corresponding to $\omega_p = 200$ cm^{-1} and $\lambda = 500$ nm shows that ROA is 10^2 more favourable than VCD in this particular case. Essentially this is due to visible light being used in the former and IR radiation in the latter.

1.2.4. Analysis of ROA and VCD: A Simple Chiral Two-group Structure

In this section only the most basic arguments of the two-group model are used to provide justification of the statements made in the previous section. A more detailed and general analysis is given by Barron (1982). If for ROA and VCD the approximations are made that

$$\text{ROA intensity} \sim \epsilon_{\beta\gamma\delta} R_{12,\gamma} (\partial\alpha_{1,\alpha\beta} / \partial\Delta\varphi_1)_0 (\partial\alpha_{2,\delta\alpha} / \partial\Delta\varphi_2)_0 \quad (1.4.1)$$

and that

$$\text{VCD intensity} \sim \varepsilon_{\alpha\beta\gamma} \mathbf{R}_{12\beta} (\partial\mu_{1\alpha} / \partial\Delta\phi_1)_0 (\partial\mu_{2\gamma} / \partial\Delta\phi_2)_0 \quad (1.4.2)$$

then the following analysis is obtained. Figure 1.4 illustrates a simple chiral two-group structure that is used to analyse the ideal modes described in the previous section. Here, $\mathbf{u}_{1\alpha} (\equiv \mathbf{u}_1)$ and $\mathbf{u}_{2\alpha} (\equiv \mathbf{u}_2)$ are unit vectors along the bond axes at any instant during a vibrational excursion, and $\mathbf{R}_{21\alpha} (\equiv \mathbf{R}_{21})$ is the inter-nuclear distance between groups two and one. This arrangement is similar to that encountered in H_2O_2 (Barron, 1982). In addition, the unit vectors $\mathbf{u}_{1\alpha}$ and $\mathbf{u}_{2\alpha}$ may be written as

$$\mathbf{u}_{1\alpha} = I_\alpha \cos\Delta\phi_1 - K_\alpha \sin\Delta\phi_1 \quad (1.5.1)$$

and

$$\mathbf{u}_{2\alpha} = I_\alpha \cos\theta \cos\Delta\phi_2 + J_\alpha \sin\theta \cos\Delta\phi_2 + K_\alpha \sin\Delta\phi_2 \quad (1.5.2)$$

where $I_\alpha (\equiv \mathbf{I})$, $J_\alpha (\equiv \mathbf{J})$ and $K_\alpha (\equiv \mathbf{K})$ are unit vectors along the internal molecular axes X , Y and Z in Figure 1.4. Since $\mu_{i\alpha} = \mu_i \mathbf{u}_{i\alpha}$ differentiation of this equation yields, for VCD

$$(\partial\mu_{i\alpha} / \partial\Delta\phi_i)_0 = \mu_i (\partial\mathbf{u}_{i\alpha} / \partial\Delta\phi_i)_0 + \mathbf{u}_{i\alpha} (\partial\mu_i / \partial\Delta\phi_i)_0 \quad (1.6.1)$$

Assuming that $(\partial\mu_i / \partial\Delta\phi_i)_0 = 0$, equation 1.6.1 becomes

$$(\partial\mu_{i\alpha} / \partial\Delta\phi_i)_0 = \mu_i (\partial\mathbf{u}_{i\alpha} / \partial\Delta\phi_i)_0 \quad (1.6.2)$$

and for $i = 1$ and 2 the following results are obtained

$$(\partial\mathbf{u}_{1\alpha} / \partial\Delta\phi_1)_0 = (-I_\alpha \sin\Delta\phi_1 - K_\alpha \cos\Delta\phi_1)_0 = -K_\alpha \quad (1.7.1)$$

and

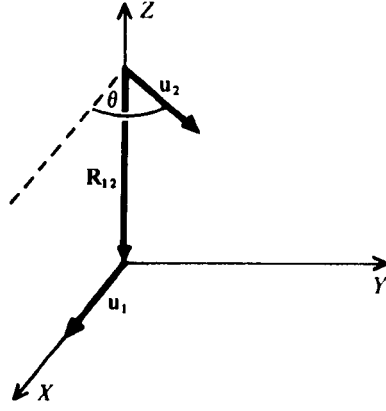


Figure 1.4 A Simple Chiral Two-group Structure. (Reproduced from Barron, 1982.)

$$(\partial u_{2\alpha} / \partial \Delta \phi_2)_0 = (-I_\gamma \cos \theta \sin \Delta \phi_2 - J_\gamma \sin \theta \sin \Delta \phi_2 + K_\gamma \cos \Delta \phi_1)_0 = K_\gamma \quad (1.7.2)$$

Therefore, if equations 1.7.1 and 1.7.2 are combined with equation 1.4.2 the following is obtained:

$$\epsilon_{\alpha\beta\gamma} R_{12\beta} (\partial \mu_{1\alpha} / \partial \Delta \phi_1)_0 (\partial \mu_{2\gamma} / \partial \Delta \phi_2)_0 = -\mu_1 \mu_2 \epsilon_{\alpha\beta\gamma} R_{21\beta} K_\alpha K_\gamma = 0 \quad (1.8.1)$$

By using the fact that $R_{21\beta} = R_{21} K_\beta$ and observing that the right-hand side of equation 1.8.1 is a vector triple product (Aklonis & Goldberg, 1977), it is easy to see that this result evaluates to zero.

A similar analysis for ROA, utilising the fact that, $\alpha_{i\alpha\beta} = \alpha_i u_{i\alpha} u_{i\beta}$, reveals after differentiation that

$$(\partial \alpha_{i\alpha\beta} / \partial \Delta \phi_i)_0 = \alpha_i (\partial (u_{i\alpha} u_{i\beta}) / \partial \Delta \phi_i)_0 + (u_{i\alpha} u_{i\beta}) (\partial \alpha_i / \partial \Delta \phi_i)_0 \quad (1.9.1)$$

Assuming that $(\partial \alpha_i / \partial \Delta \phi_i)_0 = 0$, equation 1.9.1 becomes

$$(\partial \alpha_{i\alpha\beta} / \partial \Delta \phi_i)_0 = \alpha_i (\partial (u_{i\alpha} u_{i\beta}) / \partial \Delta \phi_i)_0 \quad (1.9.2)$$

and for $i = 1$ and using equation 1.5.1 the partial derivative on the right-hand side of equation 1.9.2 gives

$$(\partial(u_{1_\alpha} u_{1_\beta}) / \partial \Delta \phi_1)_0 = -(I_\alpha K_\beta + K_\alpha I_\beta) \quad (1.10.1)$$

For $i = 2$, with equation 1.5.2 the partial derivative now becomes

$$(\partial(u_{2_\alpha} u_{2_\beta}) / \partial \Delta \phi_2)_0 = (I_\alpha K_\beta + K_\alpha I_\beta) \cos \theta + (J_\alpha K_\beta + K_\alpha J_\beta) \sin \theta \quad (1.10.2)$$

Proceeding in a similar manner to the VCD example, the results of equations 1.10.1 and 1.10.2 may be combined with equation 1.4.1 to give

$$(\partial \alpha_{1_\alpha \beta} / \partial \Delta \phi_1)_0 \epsilon_{\beta \gamma \delta} R_{21_\gamma} (\partial \alpha_{2_\alpha \beta} / \partial \Delta \phi_2)_0 = \alpha_1 \alpha_2 R_{21} \sin \theta \neq 0 \quad (1.11.1)$$

Equation 1.11.1 is clearly non-zero except when $\theta = 0^\circ$ or 180° , with both of these angles corresponding to an eclipsed and staggered achiral conformation, respectively.

1.3. Discussion

In the early days of ROA its application to biochemical problems was hindered by insufficient instrument sensitivity (Keiderling, 1996; Barron *et al.*, 1996). The ROA signal is three to four orders of magnitude weaker than the parent Raman signal and was undetectable in the conventional 90° scattering geometry, employed at the time, from biomolecules in aqueous solutions (Barron & Hecht, 1994; Barron *et al.*, 1996; Hecht *et al.*, 1999). The current setup utilises a back scattering geometry, discussed in detail in chapter three. It is now possible to run routinely the most complex of biopolymers, assuming they are not birefringent (see chapter three) or coloured. Coloured samples are not used since absorption will cause local heating of samples and will cause sample photodegradation (Nafie & Zimba, 1987). Also, it will be shown later (in chapter two) that different strategies for measuring ROA can be used (Hecht *et al.*, 1999). Over the past ten years or so novel notch filters and diffraction gratings, based on volume holographic technology, have been developed, which have brought ROA to the forefront of biochemical spectroscopic studies (Barron & Hecht, 1994; Barron *et al.*, 1996; Hecht *et al.*, 1999).

Proteins, glycoproteins, nucleic acids, viruses, and carbohydrates have all now

been studied by ROA (Barron & Hecht, 1994; Barron *et al.*, 1996; Barron *et al.*, in press). Since the Raman cross section of water is low (water makes only very weak contributions to the Raman spectrum), biopolymers in their natural environment are made accessible to ROA studies. Non-native, reduced and molten globule states (still in water) are also relatively easy to investigate (Barron *et al.*, in press). This means that natural solvation effects, especially in proteins, can be explored by ROA and compared with the solvent sensitivity of other techniques such as Raman and CD with recent results suggesting that ROA is hypersensitive to the hydration of protein secondary structures (Barron *et al.*, in press). Also, the ROA spectrum can be measured over a greater spectral range than VCD, namely from $\sim 200\text{ cm}^{-1}$ to 1800 cm^{-1} (Barron & Hecht, 1994; Barron *et al.*, 1996), although in practice the low wavenumber limit is $\sim 600\text{ cm}^{-1}$, due to interference problems from the Rayleigh wing (Bell *et al.*, 1993). In conclusion, we can see that ROA has the potential for obtaining structural information that is superior to almost any other “non-pictorial” spectroscopic method (i.e. for techniques that do not give information at atomic resolution).

Chapter 2

Theory of ROA

In developing a theory of ROA either a semi-classical or a quantum field theory approach can be used (Barron, 1982). The semi-classical approach, in which molecules are regarded as quantum objects and the radiation field as a classical unquantised perturbing field, is used here. Quantum field theory gives equivalent results but is more complicated.

The electric field vector of a plane wave light beam is given, in a complex notation, by

$$\mathbf{E} = \mathbf{E}^{(0)} \exp[-i\omega (t - \mathbf{n} \cdot \mathbf{r} / c)] \quad (2.1.1)$$

where $E^{(0)}$ is the amplitude, c is the velocity of light, \mathbf{n} is a propagation vector related to the wavevector $\mathbf{\kappa} (= \mathbf{n} \omega / c)$, with magnitude equal to the refractive index n in the direction of propagation and $\omega (= 2\pi\nu / \lambda)$ is the angular frequency, where $v = c / n$ is the velocity of the light wave in a medium. Note that in free space \mathbf{n} is a unit propagation vector (Barron, 1982).

However, the phenomenon of ROA is dependent on circularly polarised light. In this instance the light field is viewed as a superposition of two orthogonal linearly polarised plane waves with equal amplitudes, and with their phases offset by $\pm 90^\circ$, corresponding to right- and left-circular polarisation. The convention for giving the sense of circular polarisation states is that when right- or left-circularly polarised light is received by an observer, the electric field vector is seen to rotate clockwise or anticlockwise, respectively (Barron, 1982).

2.1. Molecular Property Tensors

When a molecule interacts with a radiation field, bound charges are set in motion that can be related to oscillating electric and magnetic multipole moments in

the molecule. Time dependent perturbation theory is used to derive quantum mechanical expressions for the molecular property tensors that appear in the observables of optical activity experiments. The interaction Hamiltonian and subsequent derivation of the dynamic field induced molecular property tensors are given by Barron (1982). Many molecular property tensors exist but only three are required to generate ROA. These are the electric dipole–electric dipole polarisability tensor, $\alpha_{\alpha\beta}$, familiar from conventional Raman spectroscopy, the electric dipole–magnetic dipole optical activity tensor, $G'_{\alpha\beta}$, and the electric dipole–electric quadrupole optical activity tensor, $A_{\alpha,\beta\gamma}$. These tensors are combined as products of the form $\alpha G'$ and αA to give expressions for the generation of ROA. Strictly speaking, this is only true for Rayleigh optical activity; ROA requires the use of vibrational transition tensors such as $(\alpha_{\alpha\beta})_{mn}$ (where n and m are labels for the initial and final vibrational states), along with similar expressions for the optical activity tensors. Note that throughout the rest of this chapter Cartesian tensor notation is used, in which the Einstein summation convention for repeated Greek suffices is implied.

The quantum-mechanical expressions for these molecular property tensors are given as (Barron, 1982)

$$\alpha_{\alpha\beta} = \frac{2}{\hbar} \sum_{j \neq n} \frac{\omega_{jn}}{\omega_{jn}^2 - \omega^2} \text{Re}(\langle n | \hat{\mu}_\alpha | j \rangle \langle j | \hat{\mu}_\beta | n \rangle) \quad (2.2.1)$$

$$G'_{\alpha\beta} = -\frac{2}{\hbar} \sum_{j \neq n} \frac{\omega}{\omega_{jn}^2 - \omega^2} \text{Im}(\langle n | \hat{\mu}_\alpha | j \rangle \langle j | \hat{\mu}_\beta | n \rangle) \quad (2.2.2)$$

$$A_{\alpha,\beta\gamma} = \frac{2}{\hbar} \sum_{j \neq n} \frac{\omega_{jn}}{\omega_{jn}^2 - \omega^2} \text{Re}(\langle n | \hat{\mu}_\alpha | j \rangle \langle j | \hat{\Theta}_{\beta\gamma} | n \rangle) \quad (2.2.3)$$

where $\hbar = h / 2\pi$ and $\omega_{jn} = \omega_j - \omega_n$ is the angular frequency separation between the initial and virtual intermediate states $|n\rangle$ and $|j\rangle$ of the molecule. Equation 2.2.1 gives the quantum mechanical expression for the polarisability, which is chiefly associated with refraction and light scattering phenomena; equation 2.2.2 gives that for the optical activity tensor responsible for conventional optical rotation in fluids through its trace

$G'_{\alpha\alpha}$, and equation 2.2.3 gives that for the optical activity tensor making additional contributions to optical rotation in oriented samples. In isotropic samples the $A_{\alpha,\beta\gamma}$ tensor averages to zero and makes no contribution to optical rotation or CD experiments (Barron & Hecht, 1994). For ROA in isotropic samples the situation is different with roughly equal contributions being made by $G'_{\alpha\beta}$ and $A_{\alpha,\beta\gamma}$ (Barron, 1982).

In these equations the electric dipole, magnetic dipole and the traceless electric quadrupole moment operators (in SI units) are given by (Barron, 1982)

$$\mu_{\alpha} = \sum_i e_i r_{i\alpha} \quad (2.3.1)$$

$$m_{\alpha} = \sum_i \frac{e_i}{2m_i} \epsilon_{\alpha\beta\gamma} r_{i\beta} p_{i\gamma} \quad (2.3.2)$$

$$\Theta_{\alpha\beta} = \frac{1}{2} \sum_i e_i (3r_{i\alpha} r_{i\beta} - r_i^2 \delta_{\alpha\beta}) \quad (2.3.3)$$

where particle i has position vector \mathbf{r}_i , mass m_i , charge e_i and linear momentum \mathbf{p}_i . The Kronecker delta, $\delta_{\alpha\beta}$, has values of +1 for $\alpha = \beta$ and 0 for $\alpha \neq \beta$; the Levi-Civita alternating pseudo-tensor, $\epsilon_{\alpha\beta\gamma}$, has values of +1 for an even transposition of suffices, -1 for an odd transposition of suffices and 0 if any two suffices are the same. In equation 2.3.3 above, the traceless definition for $\Theta_{\alpha\beta}$ is used (i.e. $\Theta_{\alpha\alpha} = \Theta_{xx} + \Theta_{yy} + \Theta_{zz} = 0$). Note that the α component of the vector cross product $\mathbf{r} \times \mathbf{p}$ is denoted by $\epsilon_{\alpha\beta\gamma} r_{\beta} p_{\gamma}$. Also $\delta_{\alpha\beta}$ and $\epsilon_{\alpha\beta\gamma}$ behave as polar and axial tensors (Barron, 1982), respectively. Polar and axial tensors of the same rank under spatial inversion have opposite behaviour. For example, μ_{α} is a first-rank polar tensor and changes sign whereas m_{α} is a first-rank axial tensor and so does not; and $\Theta_{\alpha\beta}$ is a second-rank polar tensor and does not change sign whereas $G'_{\alpha\beta}$ is a second-rank axial tensor and does change sign. (See equations 4.2.32a & b of Barron (1982) for details, p 156.)

2.2. ROA Observables

The strategies for measuring ROA have been, in recent years, developed into three distinct forms. In Glasgow we use the incident circular polarisation (ICP) strategy in which ROA is measured as a differential of scattered light intensities in right- and left-circularly polarised incident light (Hecht *et al.*, 1992a). A second form of ROA known as scattered circular polarisation (SCP), developed in the late 1980s (Spencer *et al.*, 1988; Hecht *et al.*, 1991; Hecht *et al.*, 1992b) measures the difference in the intensities of right- and left-circular polarisation in the scattered radiation associated with linear polarised incident light. A third and more intricate form of ROA known as dual circular polarisation (DCP) (Nafie *et al.*, 1991; Hecht & Barron, 1990b; Nafie & Freeman, 1989; Hecht & Nafie, 1990) exists, and has two distinct forms known as DCP_I and DCP_{II}. The DCP experiment is a hybrid form of the ICP and the SCP modulation strategies, and is measured when the incident radiation is switched between right- and left-circular polarised light and the differential in the intensities of the right- and left-circular polarised scattered light is measured. The DCP_I strategy, also known as in-phase DCP, takes the left-circularly polarised component of the scattered intensity generated by the left-circularly polarised incident light and is subtracted from the right-circularly polarised component of the scattered light generated by the right-circularly scattered incident light. For DCP_{II}, also known as out-of-phase DCP, the left-circularly polarised scattered light generated from right-circularly polarised incident light is subtracted from the right-circularly polarised scattered light generated from the left-circularly polarised incident light. These distinct measurement strategies are illustrated in Figure 2.1.

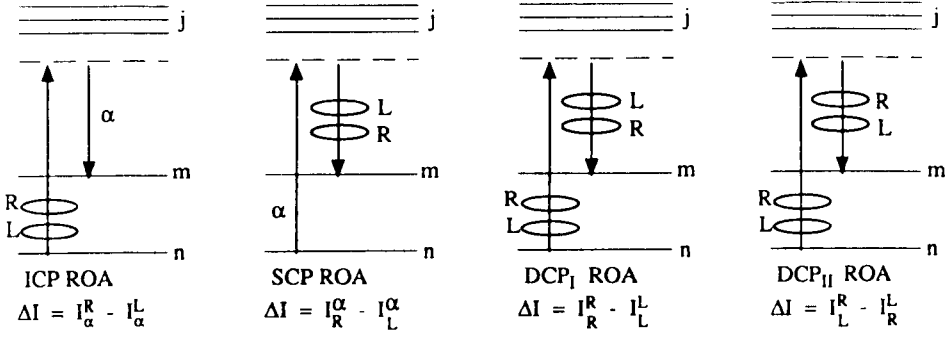


Figure. 2.1 The different modulation strategies for ROA. (Reproduced from Nafie, 1991.)

The dimensionless circular intensity differences (CIDs) associated with each of these experiments are as follows:

$$\Delta_{ICP} = (I^R - I^L)/(I^R + I^L) \quad (2.4.1)$$

$$\Delta_{SCP} = (I_R - I_L)/(I_R + I_L) \quad (2.4.2)$$

$$\Delta_{DCP_I} = (I_R^R - I_L^L)/(I_R^R + I_L^L) \quad (2.4.3)$$

$$\Delta_{DCP_{II}} = (I_L^R - I_R^L)/(I_L^R + I_R^L) \quad (2.4.4)$$

where superscripts represent the circular polarisation states of the incident light and subscripts denote the circular polarisation states of the scattered light.

In our group the ICP strategy is preferred due to its simplicity. Although the SCP approach is equivalent to ICP in the far-from-resonance approximation at transparent frequencies, experimental complexities cause a large dead time in successive measurements when using SCP; the dead time for ICP is of the order of a few tenths of a second, due to the use of an electro-optic modulator (EOM), which generates right- and left-circularly polarised incident light. SCP uses a quarter-wave plate (Nafie, 1996), which has to be moved between each signal acquisition, to detect right- and left-circularly polarised scattered light. The DCP experiment is similar in its setup to the SCP approach, with a comparable dead time but with a better signal-to-noise ratio (SNR) (Nafie, 1996).

It is emphasised that when an ICP ROA spectrum is recorded, the circular intensity difference $I^R - I^L$ and sum $I^R + I^L$ are displayed separately on the same arbitrary intensity scale (Barron, 1982). This ensures that the diagnostic value of adjacent Raman bands with conservative circular intensity difference couplets is retained. If Δ is recorded directly bands with the same $I^R - I^L$ will in general have different $I^R + I^L$ values.

2.3. Natural Raman Optical Activity

The measurement of ROA observables is achieved by selecting a well-defined experimental configuration (Barron & Hecht, 1994). This entails considerations of both the scattering geometry and the state of polarisation of the light beam. The most common scattering geometries are right-angle (90°), forward (0°) and backward (180°). In the right-angle configuration a linear Polaroid analyser may be placed in the scattered beam (illustrated in Figure 2.2) with its transmission axis either perpendicular (x) or parallel (z) to the scattering plane (yz). The polarised (x) and depolarised (z) Raman signals can be measured to give the depolarisation ratio $\rho = I_x / I_z$, which, in conventional Raman spectroscopy, gives information on the symmetry species of a molecular vibration. For ROA, deviations of $(I_x^R - I_x^L) / (I_z^R - I_z^L)$ from 2 (for right-angle only) gives a measure of the breakdown of the bond polarisability theory. (The bond polarisability theory is not discussed in any detail:- see Barron (1982) for a full discussion.)

The experimental considerations of the ROA CID observables in the last paragraph can be given tractable mathematical forms. If the polarisability-polarisability and the polarisability-optical activity tensor component products are averaged over all orientations of a scattering molecule, implying that these relations apply to isotropic samples such as neat organic liquids or aqueous solutions, the following expressions are obtained (Barron, 1982; Barron & Hecht, 1994)

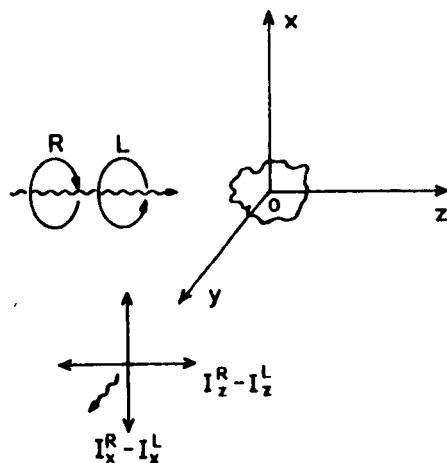


Figure 2.2 The right-angle experimental geometry for polarised ($I_x^R - I_x^L$) and depolarised ($I_z^R - I_z^L$) ROA measurements. (Reproduced from Barron & Hecht, 1994.)

$$\Delta(0^\circ) = \frac{8[45\alpha G' + \beta(G')^2 - \beta(A)^2]}{2c[45\alpha^2 + 7\beta(\alpha)^2]} \quad (2.5.1)$$

$$\Delta(180^\circ) = \frac{48[\beta(G')^2 + \frac{1}{3}\beta(A)^2]}{2c[45\alpha^2 + 7\beta(\alpha)^2]} \quad (2.5.2)$$

$$\Delta_x(90^\circ) = \frac{4[45\alpha G' + 7\beta(G') + \beta(A)^2]}{2c[45\alpha^2 + 7\beta(\alpha)^2]} \quad (2.5.3)$$

$$\Delta_z(90^\circ) = \frac{12[\beta(G')^2 - \frac{1}{3}\beta(A)^2]}{6c\beta(\alpha)^2} \quad (2.5.4)$$

where x and z refer to polarised and depolarised right-angle scattering configurations. Observe that the common factor of two is retained in the numerators and the denominators so that the intensity differences in the numerators are directly comparable. Strictly, the equations 2.5.1 to 2.5.4 apply only to Rayleigh scattering (Barron, 1982), i.e. when there is no net change in the vibrational quantum number, v . However, a glance at equations 2.5.1 to 2.5.4 will give us an idea of how the ROA signal is generated. In each of the numerators the ROA intensity stems from terms in

$\alpha G'$, $\beta(G')^2$ and $\beta(A)^2$. Each term represents the interference between the light scattered via the polarisability and one of the two optical activity tensors. Note that they also each have a sign dependence on the sense of circular polarisation. Contributions are also made by other terms such as G'^2 and A^2 but are in general about three orders of magnitude weaker than the other terms:- they are also independent of the sense of circular polarisation (Barron, 1982). The terms appearing in equations 2.5.1 to 2.5.4 are given by (Barron, 1982; Barron & Hecht, 1994)

$$\alpha = \frac{1}{3} \alpha_{\alpha\alpha} = \frac{1}{3} (\alpha_{xx} + \alpha_{yy} + \alpha_{zz}) \quad (2.6.1)$$

$$G' = \frac{1}{3} G'_{\alpha\alpha} = \frac{1}{3} (G'_{xx} + G'_{yy} + G'_{zz}) \quad (2.6.2)$$

$$\begin{aligned} \beta(\alpha)^2 &= \frac{1}{2} (3\alpha_{\alpha\beta}\alpha_{\alpha\beta} - \alpha_{\alpha\alpha}\alpha_{\beta\beta}) \\ &= \frac{1}{2} [(\alpha_{xx} - \alpha_{yy})^2 + (\alpha_{xx} - \alpha_{zz})^2 + (\alpha_{yy} - \alpha_{zz})^2 \\ &\quad + 6(\alpha_{xy}^2 + \alpha_{xz}^2 + \alpha_{yz}^2)] \end{aligned} \quad (2.6.3)$$

$$\begin{aligned} \beta(G')^2 &= \frac{1}{2} (3\alpha_{\alpha\beta}G'_{\alpha\beta} - \alpha_{\alpha\beta}G'_{\alpha\beta}) \\ &= \frac{1}{2} \{[(\alpha_{xx} - \alpha_{yy})(G'_{xx} - G'_{yy}) + (\alpha_{xx} - \alpha_{zz})(G'_{xx} - G'_{zz}) + \\ &\quad (\alpha_{yy} - \alpha_{zz})(G'_{yy} - G'_{zz}) + 3[\alpha_{xy}(G'_{xy} + G'_{yx}) \\ &\quad + \alpha_{xz}(G'_{xz} + G'_{zx}) + \alpha_{yz}(G'_{yz} + G'_{zy})]\} \end{aligned} \quad (2.6.4)$$

$$\begin{aligned} \beta(A)^2 &= \frac{1}{2} \omega \alpha_{\alpha\beta} \epsilon_{\alpha\gamma\delta} A_{\gamma\delta\beta} \\ &= \frac{1}{2} \omega [(\alpha_{yy} - \alpha_{xx}) A_{z,xy} + (\alpha_{xx} - \alpha_{zz}) A_{y,zx} + (\alpha_{zz} - \alpha_{yy}) A_{x,yz} \\ &\quad + \alpha_{xy} (A_{y,yz} - A_{y,zz} + A_{z,xx} - A_{x,xz}) + \alpha_{xz} (A_{y,zz} - A_{z,zy} + A_{x,xy} - A_{y,xx}) \\ &\quad + \alpha_{yz} (A_{z,zy} - A_{x,zz} + A_{x,yy} - A_{y,yx})] \end{aligned} \quad (2.6.5)$$

where α and G' are the isotropic invariants of the polarisability tensor and the electric-dipole magnetic dipole optical activity tensor, and $\beta(\alpha)^2$, $\beta(G')^2$ and $\beta(A)^2$ are the anisotropic invariants of the polarisability-polarisability and the polarisability-optical activity tensor component products (now referred to molecule-fixed axes).

The applicability of equations 2.5.1 to 2.5.4 to ROA is realised when the substitutions of the following transition polarisability and optical activity tensors are made to account for vibrational transitions (Barron & Hecht, 1994)

$$\alpha_{\alpha\beta} \rightarrow \langle m_v | \hat{\alpha}_{\alpha\beta}(Q_p) | n_v \rangle \quad (2.7.1)$$

$$G'_{\alpha\beta} \rightarrow \langle m_v | \hat{G}'_{\alpha\beta}(Q_p) | n_v \rangle \quad (2.7.2)$$

$$A_{\alpha,\beta\gamma} \rightarrow \langle m_v | \hat{A}_{\alpha,\beta\gamma}(Q_p) | n_v \rangle \quad (2.7.3)$$

where $|n_v\rangle$ and $|m_v\rangle$ are the initial and final vibrational states and $\hat{\alpha}_{\alpha\beta}(Q_p)$, $\hat{G}'_{\alpha\beta}(Q_p)$ and $\hat{A}_{\alpha,\beta\gamma}(Q_p)$ are the effective polarisability and the optical activity operators, respectively, that depend on the normal vibrational coordinate, Q_p .

The argument is further advanced if the transition polarisability and the optical activity tensors are expanded in normal vibrational coordinates about the equilibrium nuclear configuration, using a Taylor series expansion and assuming the Placzek approximation is applicable (i.e. at transparent frequencies), to give (Barron, 1982)

$$\langle m_v | \hat{\alpha}_{\alpha\beta}(Q_p) | n_v \rangle = (\alpha_{\alpha\beta})_0 \delta_{m_v n_v} + \sum_p (\partial \alpha_{\alpha\beta} / \partial Q_p) \langle m_v | Q_p | n_v \rangle + \dots \quad (2.8.1)$$

with similar expressions for the transition optical activity tensors. The subscript zero indicates that the terms in the series are evaluated about the equilibrium nuclear configuration. Here, the first term represents Rayleigh scattering, with the second term describing vibrational Raman scattering with the selection rule $m_v - n_v = \pm 1$, for the case of a simple harmonic oscillator. Equation 2.8.1 can now be used for the fundamental transition $I_p \leftarrow 0$ associated with the normal vibrational coordinate, Q_p , to give the following expressions for the Raman intensity and the polarisability–optical activity interference terms in the CIDs as (Barron, 1982; Barron & Hecht, 1994)

$$\langle 0 | \hat{\alpha}_{\alpha\beta} | 1_p \rangle \langle 1_p | \hat{\alpha}_{\alpha\beta} | 0 \rangle = (\hbar / 2\omega_p) (\partial \alpha_{\alpha\beta} / \partial Q_p)_0 (\partial \alpha_{\alpha\beta} / \partial Q_p)_0 \quad (2.9.1)$$

$$\langle 0 | \hat{\alpha}_{\alpha\beta} | 1_p \rangle \langle 1_p | \hat{G}'_{\alpha\beta} | 0 \rangle = (\hbar / 2\omega_p) (\partial \alpha_{\alpha\beta} / \partial Q_p)_0 (\partial G'_{\alpha\beta} / \partial Q_p)_0 \quad (2.9.2)$$

$$\langle 0 | \hat{\alpha}_{\alpha\beta} | 1_p \rangle \langle 1_p | \epsilon_{\alpha\gamma\delta} \hat{A}_{\gamma,\delta\beta} | 0 \rangle = (\hbar / 2\omega_p) (\partial \alpha_{\alpha\beta} / \partial Q_p)_0 \epsilon_{\alpha\gamma\delta} (\partial A_{\gamma,\delta\beta} / \partial Q_p)_0 \quad (2.9.3)$$

From equations 2.9.1 to 2.9.3 the symmetry requirements of Rayleigh optical activity and ROA can be extracted (Barron, 1982). For Rayleigh optical activity the same components of $\alpha_{\alpha\beta}$ and $G'_{\alpha\beta}$ must span the totally symmetric representation and, for ROA, the same components of $\alpha_{\alpha\beta}$ and $G'_{\alpha\beta}$ must span the irreducible representation of the particular normal coordinates of vibration. The chiral point groups C_n , D_n , O , T and I , which lack any improper rotation elements, are the only groups that support this requirement, in which polar and axial tensors of the same rank such as $\alpha_{\alpha\beta}$ and $G'_{\alpha\beta}$ have the same transformation characteristics. Also, the tensor $A_{\alpha,\beta\gamma}$, which does not transform like $G'_{\alpha\beta}$, is contracted to form a second rank axial tensor $\epsilon_{\alpha\gamma\delta} A_{\gamma,\delta\beta}$ that combines with $\alpha_{\alpha\beta}$ in expressions for optically active scattering to give identical transformation characteristics to $G'_{\alpha\beta}$. Thus, in principle, all Raman active vibrations are ROA active in a chiral molecule.

2.4. Backscattered ROA

Although not mentioned explicitly in chapter one, the two-group model can be formulated in terms of a bond polarisability theory (Barron, 1982). The derivation of the bond polarisability theory is not discussed here. However, a few simple results will be highlighted, which provide insight into the production of backscattered ROA signals.

In their original treatment of the two-group model Barron and Buckingham (1971) showed that, although the polarisability tensor $\alpha_{\alpha\beta}$ is independent of the choice of molecular origin, the optical activity tensors $A_{\alpha,\beta\gamma}$ and $G'_{\alpha\beta}$ are origin dependent. If the origin is moved from O to $O + a$, with a some constant vector, the following changes are made:

$$\alpha_{\alpha\beta} \rightarrow \alpha_{\alpha\beta} \quad (2.10.1)$$

$$G'_{\alpha\beta} \rightarrow G'_{\alpha\beta} - 1/2 \omega \epsilon_{\beta\gamma\delta} a_{\gamma} \alpha_{\alpha\delta} \quad (2.10.2)$$

$$A_{\alpha,\beta\gamma} \rightarrow A_{\alpha,\beta\gamma} - 3/2 a_{\beta} \alpha_{\alpha\gamma} - 3/2 a_{\gamma} \alpha_{\alpha\beta} + a_{\delta} \alpha_{\alpha\delta} \delta_{\beta\gamma} \quad (2.10.3)$$

This procedure involves the calculation of the polarisability and the optical activity tensors of the two anisotropic groups when referred to local origins. Next, the local group optical activity tensors, using the sum of the local group tensors, are calculated from a single molecular origin. Assuming cylindrical bond symmetry, the following isotropic and anisotropic ROA invariants are obtained (Barron & Hecht, 1994):

$$\alpha G' = 0 \quad (2.11.1)$$

$$\beta(G')^2 = \beta(A)^2 = -3/4 \omega \epsilon_{\beta\gamma\delta} R_{21\gamma} \alpha_{1\alpha\beta} \alpha_{2\delta\alpha} \quad (2.11.2)$$

where R_{21} is the vector between the local origins of the two anisotropic groups and $\alpha_{1\alpha\beta}$ and $\alpha_{2\delta\alpha}$ refer to the polarisability tensors on groups 1 and 2, respectively.

Consequently, there is no isotropic ROA and the anisotropic contributions from $\beta(G')^2$ and $\beta(A)^2$ are equivalent. Substituting these results into the numerators of equations 2.5.1 to 2.5.4 the dimensionless CIDs reduce to

$$\Delta(0^\circ) = 0 \quad (2.12.1)$$

$$\Delta(180^\circ) = \frac{64\beta(G')^2}{2c[45\alpha^2 + 7\beta(\alpha)^2]} \quad (2.12.2)$$

$$\Delta_x(90^\circ) = \frac{16\beta(G')^2}{c[45\alpha^2 + 7\beta(\alpha)^2]} \quad (2.12.3)$$

$$\Delta_z(90^\circ) = \frac{8\beta(G')^2}{6c\beta(\alpha)^2} \quad (2.12.4)$$

A glance at these equations reveals some important facts. Firstly, the prediction of no ROA in the forward direction is explained by invoking the need for a phase difference for the generation of ROA in the bond polarisability theory (Barron, 1982). Here, forward-scattered waves, from the two anisotropic groups, cover the same optical path length and, consequently, have the same phase when detected. This striking result was confirmed experimentally in trans-pinane, a molecule with approximate axial bonds symmetry:- its ROA spectrum in backscattering is rich and strong but has only very weak bands in forward scattering (Barron *et al.*, 1990b). Comparing the polarised right-angle with backscattering Raman and ROA measurements there is a four-fold increase in the ROA intensity and a two-fold increase in the Raman intensity over the polarised right-angle intensities (equations 2.12.2 and 2.12.3, respectively). Overall this leads to an eight-fold reduction in the time taken to acquire a given SNR in backscattering relative to polarised right-angle scattering (Barron & Hecht, 1994). Backscattering is vital for biopolymer studies as a high background, usually from solvent water, can result in the ROA signal remaining buried in the noise. Also, the analysis of backscattered ROA is simplified by the absence of terms involving $\alpha G'$ (Barron & Hecht, 1994).

It was noted in an earlier section that in-phase DCP₁ (for the backscattering configuration) has a superior SNR to SCP and ICP ROA (Nafie *et al.*, 1991; Hecht & Barron, 1990b; Nafie & Freeman, 1989; Hecht & Nafie, 1990). This can be seen when equation 2.12.2 is compared with

$$\Delta^{DCP_1}(180^\circ) = \frac{64\beta(G')^2}{12c\beta(\alpha)^2} \quad (2.13.1)$$

Clearly, the overall Δ value for DCP₁ is greater than the corresponding Δ value for ICP ROA in the backscattering configuration.

2.5. Ab-initio ROA Calculations

The use of ROA, in particular ICP ROA, for the interpretation of protein

spectra is still in its infancy. All of the interpretive work up to the time of writing this thesis has been based on making correlations between protein crystallographic data and experimental results. In recent years the work of Polavarapu (1990, 1998), which is based on the Placzek polarisability approximation, has produced some interesting, although limited, results on the ROA intensity calculations of dipeptides such as alanyl-alanine. Current ab-initio calculations are restricted to about one hundred electrons (Wilson, 1996; Polavarapu, 1998). In future studies the use of ab-initio methods may be useful for making biopolymer ROA assignments. A brief discussion on the ab-initio theory developed by Polavarapu and others is therefore justified.

In this method the polarisability and the optical activity derivatives $(\partial\alpha_{\alpha\beta}/\partial Q_p)_0$, $(\partial G'_{\alpha\beta}/\partial Q_p)_0$ and $(\partial A_{\alpha,\beta\gamma}/\partial Q_p)_0$ are used to calculate the ROA intensities. A detailed derivation of these derivatives is given by Polavarapu (1998). The required derivatives are then calculated numerically to evaluate $\alpha_{\alpha\beta}$, $(1/\omega)G'_{\alpha\beta}$ and $A_{\alpha,\beta\gamma}$ in their static limits (Amos, 1982) at the equilibrium geometry and at the geometries displaced by 0.005 Å along each atomic coordinate, which is facilitated by the quantum chemistry program, CADPAC (Amos & Rice, 1987). As well as alanyl-alanine, ROA spectra of other molecules such as methyl oxirane, tartaric acid, L-alanine, methylcyclohexanone and methylcyclopentanone have been computed in this way (Bose *et al.*, 1990; Barron *et al.*, 1992; Barron *et al.*, 1991; Polavarapu *et al.*, 1993b; Polavarapu *et al.*, 1993a). These studies show that the best results are obtained for normal modes that most directly sample the chirality of a molecular framework, such as C-C or C-O stretching coordinates found between ~ 500 and 1100 cm⁻¹. However, scaling factors, to match the theoretical and experimental frequencies, are required. Also, calculated and experimental intensities often do not correlate, but the matching of signed bands is usually excellent, and gives a method of determining absolute configurations.

Chapter 3

Instrumentation

The application of ROA to biomolecules has developed symbiotically with a concurrent development in instrument technology. A novel, yet familiar story of accident and coincidence, tells the tale of the current state-of-the-art in biopolymer ROA spectroscopy. Developments in the defence industry on the use of holographic head-up display visors for fighter pilots has led to devices such as volume holographic notch filters and diffraction gratings which are key elements of the current design of the backscattering ICP Glasgow University ROA spectrometers #3 and #4 (GUROAS3/4). The challenge of building and running a ROA spectrometer in the past was non-trivial, as ROA is such a weak phenomenon (Barron & Hecht, 1994); but thanks to new technology like this ROA measurements are now routinely performed (Hecht *et al.*, 1999).

3.1. Optical Layout

In Figure 3.1 the familiar ‘mirror with a hole’ strategy is used, as is common for normal Raman studies that utilise a backscattering geometry (Hecht *et al.*, 1992a). The design and construction of the GUROAS spectrometer is the work of L. Hecht and L. D. Barron. The exciting light source for our experiment is a single line operated continuous-wave argon ion laser (Spectra-Physics Model Stabilite 2017-05S) with a beam diameter of ~ 1.5 mm, and operating at an exciting wavelength of 514.5 nm (green light). The choice of this wavelength is a compromise: blue light would increase the Raman scattered intensity due to the $1/\lambda^4$ Rayleigh law, but the fluorescence problems associated with Raman studies of biopolymers are enhanced in this region of the electromagnetic spectrum. Fluorescence is minimised with red excitation, but the use of red light is impracticable owing to a lack of sufficient Raman intensity for current ROA studies (Barron & Hecht, 1994).

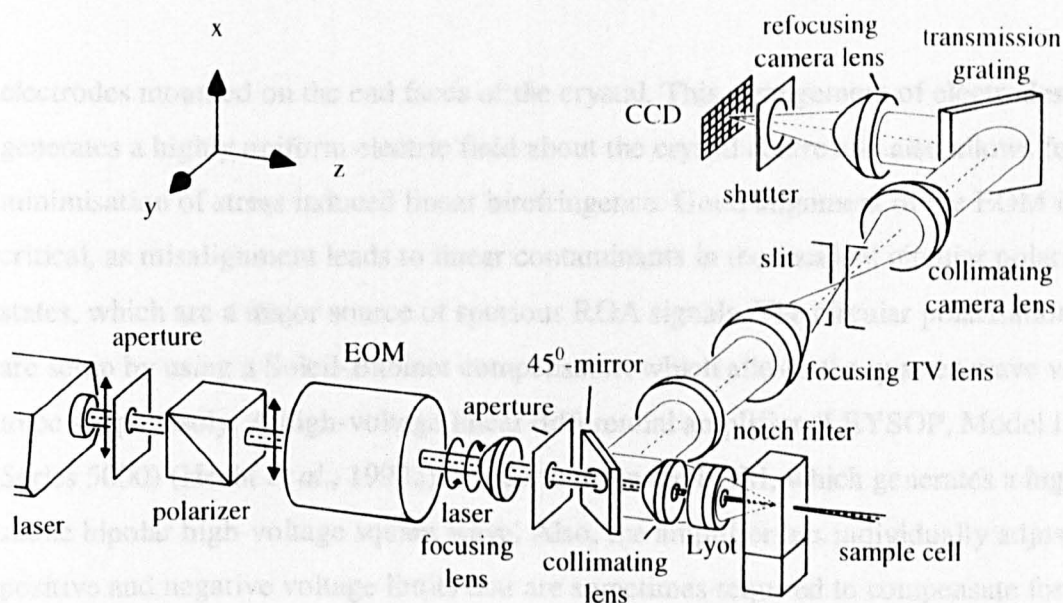


Figure 3.1 Schematic optical layout of the backscattering ROA spectrometer GUROAS3/4. (Reproduced from Hecht *et al.*, 1999.)

After exiting the laser cavity the light source passes through a circular light stop of diameter ~ 2 mm, which is mounted directly in front of the laser head, to suppress any plasma background radiation. Next the laser light is passed through an air-spaced calcite Glan-Taylor prism polariser (LEYSOP, Model GT12) (Hecht *et al.*, 1992a) with a 12×12 mm² cross-section, which augments the degree of linear (x) polarisation of the laser light. A customised electro-optic modulator (EOM) with a clear aperture of 12 mm (LEYSOP) is next encountered and, by applying predetermined quarter-wave voltages across its electrodes, at a particular wavelength, the EOM will act as a quarter-wave plate retarder, converting linear polarised laser light into circular polarised light, within a high degree of accuracy. Within the housing of the EOM a z-cut potassium dideuteriumphosphate (KD*P) crystal ($12 \times 12 \times 25$ mm³) is used to modulate between right- and left-circularly polarised light. This is achieved by passing a high voltage (~ 3000 V) and also by reversing the direction of the electric field across the crystal. To eliminate birefringence drift in the EOM, caused by external temperature perturbations, an oven is used to maintain a near constant operating temperature of $26 \pm 0.1^\circ\text{C}$, which is controlled by a small temperature controller (LEYSOP). A longitudinal electric field is generated and maintained by two gold ring electrodes together with two end face

electrodes mounted on the end faces of the crystal. This arrangement of electrodes generates a highly uniform electric field about the crystal centre and also allows for the minimisation of stress induced linear birefringence. Good alignment of the EOM is critical, as misalignment leads to linear contaminants in the incident circular polarisation states, which are a major source of spurious ROA signals. The circular polarisation states are setup by using a Soleil-Babinet compensator, which allows the quarter-wave voltages to be set precisely. A high-voltage linear differential amplifier (LEYSOP, Model LDA 50 Series 5000) (Hecht *et al.*, 1992a) is used to drive the EOM, which generates a highly stable bipolar high-voltage square wave. Also, the amplifier has individually adjustable positive and negative voltage limits that are sometimes required to compensate for birefringence effects in different samples. Modulation frequencies are controlled via a parallel interface connected to a computer. A custom-built electronics module, also connected to the parallel interface, ensures that the modulation of polarised light begins with the correct phase as well as resetting the voltage to zero between successive cycles.

After leaving the EOM the laser light is directly focused into a sample by a 150 mm focal length, 25.4 mm diameter, plano-convex synthetic fused-silica lens (Melles Griot, Model 01 LPQ 013). A small aperture ~ 1 mm is then used to reject side beams of laser light, from the optical train (e.g. from the glass panel over the laser exit), before the laser light passes through an optical block arrangement consisting of a 45° mirror, a collimating lens and a Lyot depolariser (Hecht *et al.*, 1992a). Each of these components has a small hole ~ 2 mm in diameter drilled through the middle of each element, allowing the light beam to pass through unhindered, assuming these components are aligned correctly.

After exiting the optical block the beam passes through a sample cell. The sample cell is a $6 \times 6 \times 25$ mm rectangular quartz (Spectrosil) microfluorescence cell (Optiglass) with an inner compartment path length of 5 mm, supporting a maximum sample volume of ~ 300 μ l. In practice volumes of ~ 150 μ l are used. Annealing of the sample cell at $\sim 1200^\circ\text{C}$ is required to remove any residual birefringence in the glass to prevent artefacts. An optical mount (Newport) supports the sample cell. The mount is fitted with screws that allow translational freedom in the x-, y- and z-directions. This translational freedom is vital for the correct alignment of the sample cell, as incorrectly back-reflected light

from the cell walls will enter into the collection optics of the spectrograph, resulting in baseline slopes and other artefact problems. Also, the location of the focused beam within the sample cell is important. If the beam is not focused approximately in the centre of the compartment, backscattered Raman light (around the negative z -direction) may be cut off by the side walls of the sample cell (Hecht & Barron, 1995).

The backscattered cone of Raman light, from the sample, passes back through the optical block where a Lyot depolariser (Mochizuki, 1984; Takada *et al.*, 1988; Kliger *et al.*, 1990) scrambles residual linear polarised light. Two calcite plates of different thickness in a ratio of 2:1 (Loeber, 1982; Billings, 1951; Clark & Gringer, 1971), with the thick face nearest to the sample to allow the cone of light to sample both plates, are cut from the same X-cut of a raw calcite crystal to produce the Lyot depolariser. The calcite plates are held together with optical cement with their optical axes oriented at 45° with respect to each other. Antireflection coatings are applied to the front and back faces of the Lyot depolariser. The Lyot depolariser is mounted in a rotatable stage (New Focus, Model 8401), that allows it to be rotated about the z -direction so that the optimum depolarisation position is found *via* a trial and error procedure.

Collimation of the depolarised Raman light is achieved by a 25.4 mm diameter fused silica spherical plano-convex lens (CVI Laser, Model PLCX-25.4-12.9-UV) with broad-band antireflection coating. After collimation the light is deflected through an angle of 90° by a 45° oriented mirror of dimensions 35.4×25.0 mm (Optiglass) with an enhanced silver coating (Balzers, Siflex), which exhibits a minimal difference in the reflection coefficients for horizontal and vertical linearly polarised visible light of 0.5-0.6% (Hecht *et al.*, 1992a; Clark & Gringer, 1971). This is superior to the aluminium-coated mirrors with a 5-6% (Hecht *et al.*, 1992a; Clark & Gringer, 1971) reflection coefficient difference, meaning that the silver-coated mirror represents a ten-fold improvement in the suppression of polarisation artefacts (Hecht *et al.*, 1992a).

After the light is deflected through 90° a mixture of Rayleigh and Raman light is present. Before entering the spectrograph the Rayleigh light is rejected by a holographic ‘super notch-plus’ (HSNPF) type filter (Kaiser Optical Systems, Model HSNPF-2432) with a high optical density (> 8). The HSNPF is a vital component of our spectrometer, as stray Rayleigh light is detrimental to the signal-to-noise ratio and will introduce

baseline slopes into a ROA spectrum. The HSNPF is placed directly to the side of the optical block (Hecht *et al.*, 1999) where quasi-parallel light impinges on the HSNPF and a greater efficacy of Rayleigh stray light rejection is achieved (Hecht *et al.*, 1999). An illustration of the notch filter's performance is given in Figure 3.2. Notice that the Rayleigh line rejection is at 515.5 nm and that the transmission performance of the filter is $> 80\%$ at wavelengths corresponding to a typical biopolymer Raman spectrum (~ 500 to 1700 cm^{-1}). Also, the small edge width of the filter gives a sharp wavenumber cutoff at $\sim 150\text{ cm}^{-1}$ for aqueous solutions of biopolymers (Hecht *et al.*, 1999).

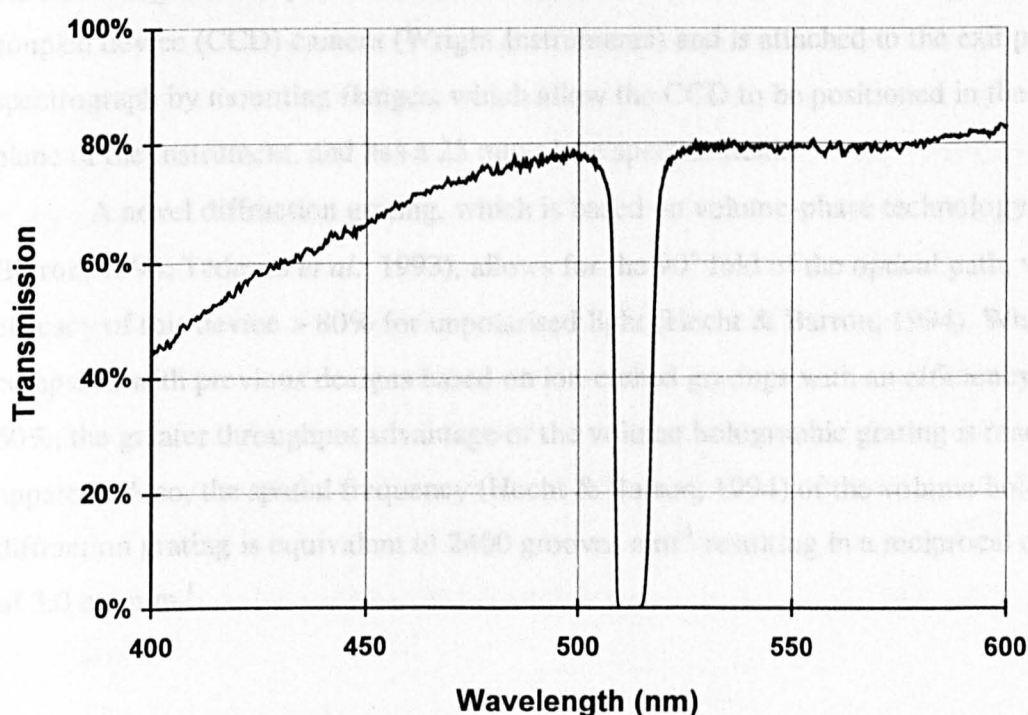


Figure 3.2 Performance of the holographic 'super notch-plus' filter. (Reproduced from Kaiser Optical Systems.)

Before entering the spectrograph the quasi-parallel light is focused by a 75 mm $f/1.4$ CCTV camera lens (Nikon) onto a 8 mm high entrance slit of a single-stage stigmatic spectrograph (Kaiser Optical System, Model Holo-spec HS- $f/2.2$), which utilises an on-axis design (Hecht & Barron, 1994). In front of this slit an electronically actuated Uniblitz shutter (Vincent associates, Model VS 25) is positioned, and is opened

at the start of a spectral run and only closes at the end of the run. The designation $f/1.4$ refers to the stop- or f-number, $f/\# = f/D$, with f the focal length and D the aperture diameter of a camera lens, which means that a smaller f-number permits more light to reach the image plane of the CCD (Kliger *et al.*, 1990). Within the spectrograph an 85 mm $f/1.4$ (Nikon) TV camera lens collimates the Raman light before the beam is dispersed through a holographic transmission grating. To maintain an $f/1.4$ cone throughout the spectrograph a refocusing lens, an identical $f/1.4$ (Nikon) TV camera lens to the one just mentioned, is placed after the grating where the refocused light finally passes through an exit port. The final component in the GUROAS3/4 setup is a charge-coupled device (CCD) camera (Wright Instruments) and is attached to the exit port of the spectrograph by mounting flanges, which allow the CCD to be positioned in the focal plane of the instrument, and has a 25 mm plane spectral field.

A novel diffraction grating, which is based on volume-phase technology (Hecht & Barron, 1994; Tedesco *et al.*, 1993), allows for the 90° fold of the optical path, with the efficacy of this device $> 80\%$ for unpolarised light (Hecht & Barron, 1994). When compared with previous designs based on ion-etched gratings with an efficiency of $\sim 60\%$, the greater throughput advantage of the volume holographic grating is readily apparent. Also, the spatial frequency (Hecht & Barron, 1994) of the volume holographic diffraction grating is equivalent to $2400 \text{ grooves mm}^{-1}$ resulting in a reciprocal dispersion of 3.0 nm mm^{-1} .

3.2. Detection System (CCD)

Towards the end of the last section the use of a CCD camera was mentioned. This device merits its own discussion. As well as the recent advances in optical technology, similar developments in detection systems have also been of great significance to the biochemical spectroscopist. CCD devices have been around for some time, since the early 1970s (Benyon *et al.*, 1980; Gudehus & Hegyi, 1985; Pemberton *et al.*, 1990), and were originally used by astronomers, as well as finding applications in the home entertainment industry for hand-held playback video recorders. By the late 1970s to early 1980s, it was

realised by spectroscopists that the CCD device is an excellent tool for the detection of low light level sources such as those found in Raman spectroscopy (Surbeck *et al.*, 1981; Epperson & Denton, 1990; Instruction Manual, Wright Instruments). A backthinned, back-illuminated CCD detection system (Hecht *et al.*, 1999) of physical dimensions $6.912 \times 27.648 \text{ mm}^2$ (EEV, Model CCD 15-11-232) and consisting of 1024 columns by 256 rows of individually spaced square pixels is used in our current setup.

The surface of the CCD is fabricated from silicon oxides (Benyon *et al.*, 1980; Surbeck *et al.*, 1981; Epperson & Denton, 1990; Instruction Manual, Wright Instruments) that are light sensitive and produce photo-induced electronic charge when impinged upon by Raman scattered light. With 514.5 laser excitation a maximum spectral coverage of $\sim 2500 \text{ cm}^{-1}$ for Stokes Raman shifts may be obtained. To achieve this coverage the CCD is placed along its long axis, parallel to the direction of dispersion of the spectrograph. Also, to achieve maximum efficiency the CCD is maintained at an operating temperature of $\sim 200\text{K}$. This is facilitated by the thermo-electric Peltier effect, allowing the detector to have a dark current (thermally induced signals) of $\sim 0.14 \text{ e}^-$ counts per pixel per second, resulting in a better SNR when compared to a CCD operating at room temperature. One final point to note is that backthinning, by an acid etching process, increases the quantum efficiency of the CCD when compared with conventional non-thinned devices (illustrated in Figure 3.3). This is easily understood by realising that in a thinned device the photo-induced charges have less distance to travel before being detected (Benyon *et al.*, 1980).

A serial pixel binning method (Tedesco *et al.*, 1993) is used to convert the two-dimensional CCD image detector into a linear array. In the binning process adjacent CCD pixels have their signals added together either on-chip or digitally following readout. On-chip binning is preferred because it reduces the total readout noise associated with the CCD output amplifier (Epperson & Denton, 1990).

Overall the net result is a highly effective detection device with a quantum efficiency in the region of $\sim 80\%$ in the 450 to 650 nm range. Less elaborate CCD designs have quantum efficiencies in the range of $\sim 20\%$ at 500 nm and rising to $\sim 50\%$ at 700 nm. Unlike photo-diode arrays no intensification of the basic signal is required, due to the generally low noise levels associated with CCDs (Benyon *et al.*, 1990; Surbeck

et al., 1981; Epperson & Denton, 1990; Instruction Manual, Wright Instruments). The performance of the CCD used in our lab is illustrated below in Figure 3.3.

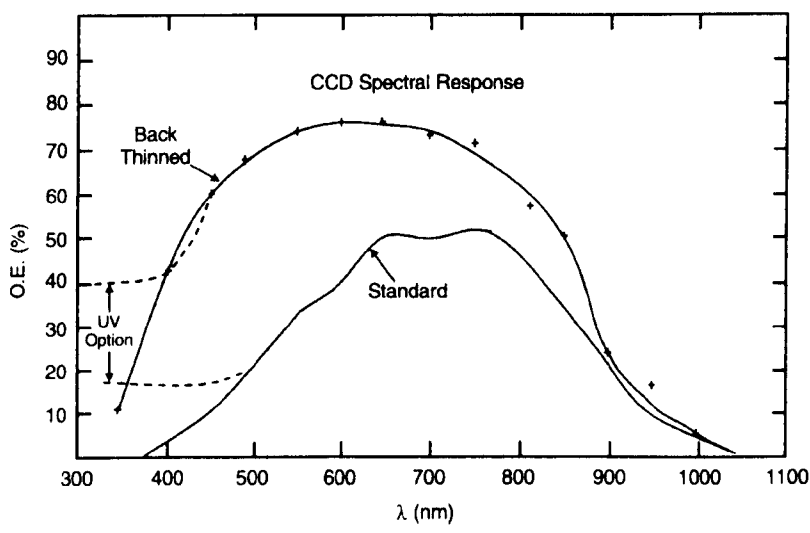


Figure 3.3 Performance of a backthinned, back-illuminated CCD camera. (Reproduced from Falkin & Vosloo, 1993.)

3.3. Computer Control

The Raman spectral acquisition in right- and left-circularly polarised incident light is controlled by a Dell personal computer fitted with a control interface to synchronise the CCD camera and the EOM. A customised version of LabCalc (Galactic Industries) is used to coordinate the processing of spectral signals. The acquisition cycle begins when the ROA electronics and the EOM are reset to produces right-circularly polarised light. The signal generated on the chip by the impinging light is then sent to a display to monitor each spectral event for any problems that may arise.

The signal is then stored in memory before the CCD is cleared of residual charge. This cycle is then repeated for left-circularly polarised light. By convention the left Raman spectrum is subtracted from the right Raman spectrum to produce a ROA spectrum. A superposition of the right and left spectra is displayed along with the ROA

spectrum in a stacked plot for comparison. Subsequently, the ROA spectrum is stored in memory, and the next cycle begins. As each cycle ends the previously stored ROA spectrum is deleted from memory after being added to the newly acquired ROA spectrum. Over a period of time ~ 5 to ~ 20 hrs the ROA signal is allowed to build up until an acceptable SNR is attained, which is judged by eye. A schematic representation of this setup is illustrated in Figure 3.4.

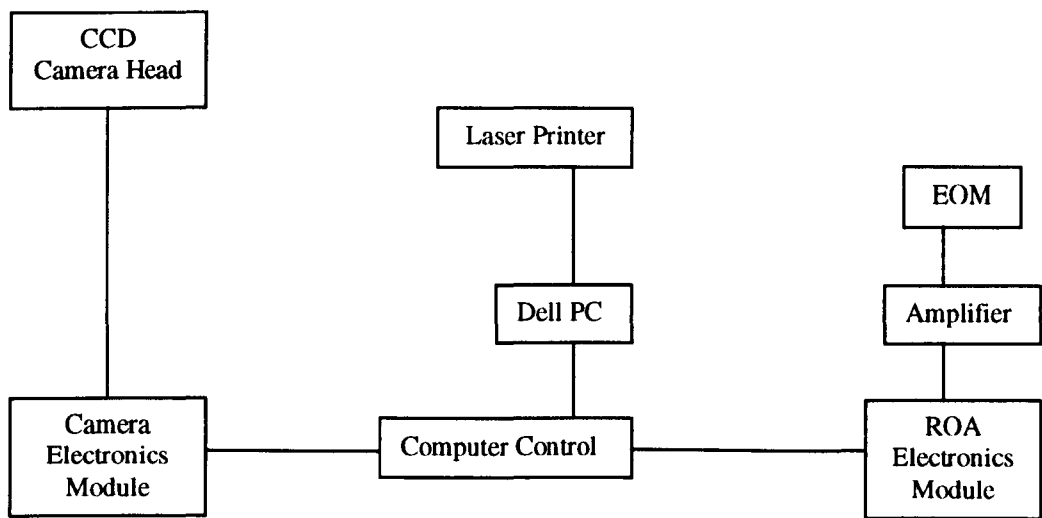


Figure 3.4 Schematic of control setup for ROA experiment.

3.4. Artefact Control

Being a weak phenomenon, the problem of artefact control is central to acquiring a ‘good’ spectrum (Barron *et al.*, 1973; Barron & Hecht, 1994). The shot noise $N_s = \sqrt{N}$ (with N the number of photons scattered) (Kliger *et al.*, 1990; Benyon *et al.*, 1980), which is a property of the collected light and not the detection system (Kliger *et al.*, 1990; Benyon *et al.*, 1980), must be a few orders of magnitude weaker than the measured ROA signal. For example, typical ROA measurements of the Δ -value are of the order $\sim 10^{-3}$ (Barron & Hecht, 1994). If $N = 10^6$ and N_s is 10^3 , then the ratio of N_s/N (shot noise /

detected photons) = 10^{-3} , as required. Sharp S-shaped signals on strongly polarised Raman bands (Barron & Hecht, 1994; Barron *et al.*, 1973; Barron & Buckingham, 1975) are one of the main artefacts encountered in a ROA experiment, with other artefact problems associated with baseline slopes and drifts. The phenomenological Stokes-Muller (SMF) formalism (Takada *et al.*, 1988; Hecht *et al.*, 1987; Gerrard & Burch, 1994) is used to provide insight into the generation of spurious signals. In the SMF a series of matrices are calculated that correspond to individual components in the spectrometer setup (Hecht *et al.*, 1987; Gerrard & Burch, 1994). Matrix parameters are varied to provide information on how best to optimise the experimental configuration of a spectrometer (Hecht *et al.*, 1987; Gerrard & Burch, 1994).

A detailed analysis (Hug, 1981; Hecht *et al.*, 1987) revealed that a significant source of ROA artefacts stem from optical components with residual linear polarisation, so that strain-free quartz components are used wherever possible. Also, within the SMF the scattering geometry of the experiment is considered. The analysis showed that the backscattering configuration has a distinct advantage over 90° scattering, as backscattered light has near-perfect rotational symmetry (Escribano, 1985; Hecht & Barron, 1990a). Consequently, the azimuthal scattering angle, which is a strong determiner of artefacts, has its effect minimised in the backscattering arrangement.

Back-reflected light, from elements in the optical train, also cause problems (Hug, 1981) by passing light back into the laser cavity, causing the degree of polarisation of the incident light to degrade. To further reduce the problem of back-reflection all the optics are antireflection coated with MgF_2 (Optiglass), wherever possible (Hug, 1981).

A more minor cause of artefacts is the optical rotation of linear polarised light contaminants, which can be minimised by using a small sample cell (Hug, 1981; Hecht & Barron, 1994). Finally, dust particles must be removed from the sample by microfiltering and centrifugation to ensure that spurious light scattering does not occur.

One final source of artefacts is the 45° mirror. If the back-reflected light contains any residual linear polarised contaminants, large polarisation artefacts are induced by the mirror (Takada *et al.*, 1988). For this reason the Lyot depolariser (Hecht & Barron, 1994) is placed in front of the mirror to scramble any residual linear polarisation. The use of the Lyot is in some ways unexpected, as the literature implies that its effectiveness depends

on the use of broadband polychromatic light (Kliger *et al.*, 1990). The Lyot's power as a retarder is a function of the different retardance that it has to differing wavelengths of visible light (Kliger *et al.*, 1990). However, backscattered Raman light (green light) is collected over a wide backscattered cone, meaning that light is collected over many azimuthal angles with differing degrees of linear polarisation, which interfere with each other resulting in pseudo-depolarisation (Kliger *et al.*, 1990). This deviation from parallel light is the key to the power of the Lyot in depolarised backscattering ROA.

A 'test' dipeptide, L-alanyl-L-alanine, is used to find the optimum orientation of the Lyot. A strongly polarised band at $\sim 880\text{ cm}^{-1}$ is seen to produce a large S-shaped signal in different orientations (Barron *et al.*, 1990a; Barron & Vrbancich, 1984). At some experimentally predetermined position, obtained by rotating the Lyot in steps of 10° to 20° the S-band will disappear or be minimised. Figure 3.5 shows an 'optimised' ROA spectrum of L-alanyl-L-alanine.

3.5. Miscellaneous

In this penultimate section, a few observations are set down on instrument performance, as a result of experience gained in the course of this project. Glycoproteins caused the main problems encountered, due to the fluorescence from samples degrading in the laser beam. As a general rule it was found that nonglycosylated proteins are well behaved and relatively easy to study.

Baseline slopes are one of the most common problems encountered with protein ROA measurements. Three main factors relating to the HSNPF can be used to minimise this problem. Namely, the tilt of the notch filter, a black spot located in the approximate centre of the filter and an aperture around the outer edge of the filter. The optimum tilt angle of the HSNPF, based on factory settings, was sometimes found to aggravate baseline problems with certain samples. This problem was minimised by making adjustments to the tilt angle by $\sim 5^\circ$. In addition, spurious signals on the periphery of the

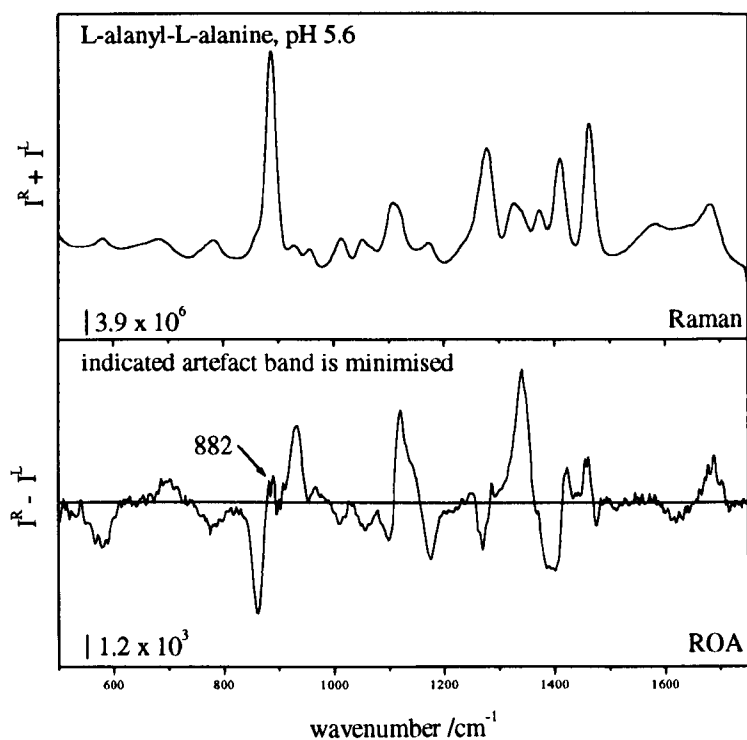


Figure 3.5 Optimised ROA spectrum of *L*-alanyl-*L*-alanine. (Reproduced from Wilson, 1996.)

notch filter, from poor collimation of light, was blocked by placing a black paper aperture around the edge of the filter. Also, a black spot is placed in the centre of the filter. A bright ring of light, caused by light clipping the hole drilled in the centre of the Lyot and also by the hole in the mirror, is usually seen impinging on the filter, which is blocked by the black spot. Adjusting the position of the spot can have a dramatic effect on the appearance and disappearance of slopes. One additional point relating to the tilt angle of the notch filter is the appearance of distortions in the α -helix backbone region at ~ 850 to 950 cm^{-1} . On some occasions the removal of sloping artefacts would lead to an enhancement in the backbone region of a spectrum, meaning that a ‘compromise’ tilt angle was used.

The appearance of calcite bands was also an occasional problem. This was observed to be a function of the sample cell position, which in turn was a function of the laser focusing lens used. A 25 mm or a 30 mm laser focusing lens was used depending on the samples studied. For glycoproteins a longer focal length was preferable. The longer focal length means that a greater collection pathlength is used over the sample and

appears to make glycoprotein studies more amenable. However, the 30 mm lens induces a calcite band $\sim 1088\text{ cm}^{-1}$, caused by back-reflected light (from the front of the sample cell) clipping the edge of the hole through the middle of the Lyot. Also, if the sample cell is moved closer to the optical block, polarised bands are boosted and are difficult to control.

To help overcome some of the problems mentioned in this section a recent enhancement to the GUROAS3/4 configuration is the use of a half-wave plate retarder. In our new setup the half-wave plate is placed between the EOM and the optical block. This arrangement means that if the half-wave plate is rotated continuously during data acquisition, the effects of linear contaminants from the EOM (the EOM does not operate perfectly in its production of right- and left-circularly polarised light) are minimised, removing one of the main causes of polarisation artefacts. Unfortunately, this improvement was implemented after the experimental part of the project was finished.

Problems were also encountered with the CCD. Occasional ‘jumps’ in glycoprotein and protein ROA spectra were seen to occur as a ROA run progressed. In the Raman spectrum this jump would appear as a separation in the right and left Raman spectra, which are displayed together, with one superposed over the other. In the Raman spectrum this jump would spread from the high wavenumber end down to the low wavenumber end of the Raman spectrum; whereas, in the ROA spectrum a sharp negative step would appear below $\sim 1100\text{ cm}^{-1}$. Static discharges were partly responsible for this problem and were associated with sudden changes in the air quality of the laboratory’s atmosphere, which was caused by people opening and closing the laboratory door. A Faraday cage was constructed to help overcome this problem. Although partially successful, occasional jumps persisted and were caused by a laser transformer pack. The CCD computer interface card had a similar effect, due to a faulty interface port. All of these problems were eventually remedied.

3.6. Sample Sources and Experimental Conditions

Samples of immunoglobulin G, porcine pepsin, ovalbumin, chicken and turkey ovomucoid, glycosylated α_1 -antiproteinase, Bowman-Birk inhibitor, invertase (grade VII from baker's yeast), Ribonuclease A and B, porcine insulin, concanavalin A, human serum albumin and bovine serum albumin, poly(L-lysine) and poly(L-glutamic acid) and ovotransferrin (substantially iron free) were all purchased from Sigma. Avidin and aldolase were purchased from Fluka, with samples of recombinant (non-glycosylated) α_1 -antiproteinase donated by D. A. Lomas of Cambridge University, metallothionein by M. Vařák of the University of Zürich and dehydroquinase (Type I) by J. R. Coggins of the University of Glasgow. All of the protein samples were dissolved in a 50 mM acetate buffer at ~ pH 5.4, except for insulin which was prepared at ~ pHs 2.2 (dimer) and 0.8 (monomer) with both of these samples obtained by adding concentrated HCl to water. Also, phosvitin was prepared at ~ pH 6 in water. The D₂O samples were made under similar conditions. The homopolypeptide samples of poly(L-lysine) and poly(L-glutamic acid) were both prepared in 50 mM glycine buffers at ~ pH 3 and 10, respectively. All of these samples were prepared at a concentration of ~ 70 mg/ml, except for α_1 -antiproteinase and dehydroquinase, which were both at a concentration of ~ 38.5 mg/ml and 30 mg/ml, respectively. The polysaccharide samples of dextran from *Leuconostoc mesenteroides* and glycogen from rabbit liver were purchased from Fluka, and had number degrees of polymerisation of ~ 300 and 10⁵, respectively. Pullulan from *Aureobasidium pullulans* and laminarin from *Laminaria digitata* were purchased from Sigma and had number degrees of polymerisation of ~ 250 and 25 to 31, respectively. All of these samples were prepared in a 50 mM Tris(hydroxymethyl)aminomethane (Tris) buffer at ~ pH 7.5. Dextran, laminarin and pullulan were prepared at a concentration of 200 mg/ml and glycogen at 50 mg/ml. The protein and homopolypeptide samples all had an acquisition time of 8 to 10 hours, except for recombinant and glycosylated α_1 -antiproteinase which both had an acquisition time of 20 hours. The polysaccharide samples all had an acquisition time of 7 hours.

Chapter 4

The Structure of Proteins and Glycoproteins

This chapter presents an overview of protein structure. As we shall see, the architecture of a protein determines its biological activity (Brandon & Tooze, 1992) in terms of specific three-dimensional folded structures that are particular to a given class of biomolecules (Thorne *et al.*, 1996). In addition, a brief description of vibrational analysis is given to provide insight into the justifications that vibrational spectroscopists use in their assignment of biopolymer Raman bands.

4.1. Protein Structural Hierarchy

A detailed structural description of a protein is multi-layered and is defined in terms of the Linderstrøm-Lang (1952) hierarchy. The ‘primary’ structure refers to a given sequence of amino acid residues in a protein’s peptide backbone linkages, with only twenty of these residues appearing in ‘naturally’ occurring proteins. All but one of these amino acids is chiral, with glycine being the exception. The ‘secondary’ structure describes the local folding of the polypeptide chain into compact structural elements such as α -helices (Pauling & Corey, 1951a) and β -sheets (Pauling & Corey, 1951b) with the relative spatial arrangement of secondary structures, which utilise long range contacts, referred to as the ‘tertiary’ structure. The association of subunit structures (e.g. the interaction between two or more folded domains) is described as the ‘quarternary’ structure.

The primary structure of a protein is often called the second genetic code (Lillo, 1997). In fact, the three-dimensional structure and function of a protein ultimately depends on this sequence. Folding studies have shown that if a protein is allowed to unfold, either by adding denaturants, varying the pH, varying the temperature or by adding reducing agents to cleave disulphide bonds, the unfolded protein (U) can be induced to refold to its

native state (N) by restoring the original native conditions of the protein (Carter, 1998). This implies that the amino acid sequence is the main source of structural information for a proteins native folded state (Anfinsen, 1973). Although forward and backward read protein sequences are not usually palindromic, theoretical lattice model studies show that a backwards read protein amino acid sequence (a retro-protein) will lead to a native-like tertiary fold close to that of a forward read protein (Olszewski *et al.*, 1996). Some differences in secondary structure and hydrophobic side chain bonding patterns are observed, but are largely conserved, with turn regions being most likely to be in disagreement for real and retro-proteins (Olszewski, 1996).

Protein secondary structures fall into three main classes: α -helical, β -sheet (antiparallel and parallel) and, turn or hairpin structures. Table 4.1 gives the Ramachandran (ϕ, ψ) angles (Carter, 1998; Ramachandran *et al.*, 1963; Muñoz & Serrano, 1994) of these structural elements.

Table 4.1 The Ramachandran angles for the most common secondary structure elements.

Structure	ϕ	ψ
α -helix	-60°	-45°
β -sheet (antiparallel) [‡]	$-142^\circ \pm 13^\circ$	$145^\circ \pm 13^\circ$
β -sheet (parallel) [‡]	$-118^\circ \pm 13^\circ$	$112^\circ \pm 13^\circ$
PPII [‡]	Between -170° and -120°	Between 110° and 170°
β -turn (type I) [*]	$\phi_1 = -54^\circ$ and $\phi_2 = -95^\circ$	$\psi_1 = -11^\circ$ and $\psi_2 = 9^\circ$
β -turn (type II) [*]	$\phi_1 = -48^\circ$ and $\phi_2 = -88^\circ$	$\psi_1 = 137^\circ$ and $\psi_2 = -20^\circ$
β -turn (type III = 3_{10} -helix) [*]	$\phi_1 = -45^\circ$ and $\phi_2 = -48^\circ$	$\psi_1 = -36^\circ$ and $\psi_2 = -36^\circ$

^{*}Although β -turns have four residues only the central two residues, $i + 1$ and $i + 2$, are required to specify their conformation (Muñoz & Serrano, 1994). [‡] Average (ϕ, ψ) angles for alanine (Muñoz & Serrano, 1994).

High resolution X-ray data is used (Carter, 1998) along with theoretical calculations (Scully & Hermans, 1994) to obtain precise values for the Ramachandran (ϕ, ψ) angles seen in table 4.1. If all of the residues in a protein have their (ϕ, ψ) angles plotted in a two-dimensional graph, a Ramachandran plot is obtained (Ramachandran *et al.*, 1963). A series of contour lines, representing the stereochemically accessible regions of a proteins structure, imply a thermodynamic tendency for residues along a polypeptide chain to collect in potential energy minima. Broadly speaking, most plots show a propensity for residues within proteins to populate regions associated with α -helices and β -sheets (Ramachandran *et al.*, 1963).

Two important reasons for secondary structure formation are the burial of backbone carbonyl and amide groups, in the form of a hydrogen bond network, and the formation of a surface area of side chains (Carter, 1998). The formation of hydrogen bonds is favourable as they allow the Ramachandran (ϕ, ψ) angles to lie within the optimum regions for α and β secondary structure (Ramachandran *et al.*, 1963). Side chain interactions present a large surface area for van der Waals contacts between complementary structures (Chan & Dill, 1990). In addition, molecular dynamics calculations demonstrate that side chain–side chain interactions are significant in the formation of secondary structure (Janin *et al.*, 1978). The hydrophobic effect is often cited as the primary reason for protein folding (Privalov, 1992) with two opposing entropic factors dominating the argument. When a polypeptide folds, a negative, unfavourable entropic effect destabilises the polypeptide fold (Privalov, 1992); however, a positive stabilising entropy contribution is also made, as a result of the dehydration of non-polar residues that become buried in the core of the polypeptide fold. Both of these factors compete with each other (Privalov, 1992). The exact nature of the hydrophobic effect is still not fully understood (Privalov, 1992). The results of an experiment must be interpreted by theoretical models of protein folding which assume that the polypeptide chain starts of from a fully unfolded state (see chapter 7). Also, it is assumed that all of the residues in the polypeptide chain do not interact with each other (Privalov, 1992). Such ideal states, however, do not exist in reality.

Secondary structure elements are connected by loops and turns (Chou & Fasman,

1979; Rose *et al.*, 1985). Without turn structures proteins would be unable to form compact globular shapes and would remain fibrous (Carter, 1998). The β -turn structure is characterised by four residues, with the central residues ($i + 1$ and $i + 2$) being the determiners of the type of turn formed, depending on which (ϕ, ψ) angles are selected (see table 4.1). Other turn structures can be defined such as zeta loops, which are essentially disordered strands ~ 14 residues long (Artymiuk *et al.*, 1994). Gamma-turns (classic and inverse) are also important and have been assigned spectroscopically by Krimm (Krimm & Bandekar, 1986). On a similar theme the 3_{10} -helix (c.f. the α -helix which is a 3.63_{13} -helix) is seen to be related to β -turn structure (it is actually a type III turn) and is often excluded from traditional classifications (Hutchinson & Thornton, 1994). Thermodynamically 3_{10} -helix is much less stable than α -helix (Carter, 1998), due to its greater rigidity (sharper turn: 3_{10} versus 3.63_{13}). Typically, most 3_{10} -helices are found at the ends of α -helices, and are only one turn in length (Karpen *et al.*, 1992) where they effect a change in direction. The study of protein secondary structure is far too intricate to discuss in any detail, with other matters to consider such as the right-handed twisted nature of β -sheets, β -bulges, β -helices, supersecondary structure and numerous other structural elements (Branden & Tooze, 1992; Carter, 1998).

One final consideration in the study of secondary structure is the validity of the X-ray data used to determine these structural parameters. It is assumed that high-resolution data gives the same representation of intra/intermolecular hydrogen bonds as low-resolution data (Carter, 1998). An alternative, and not impossible interpretation, is that the crystal properties (diffraction properties) of high-resolution structures are different to low-resolution structures (Carter, 1998). Clearly, care must be taken in the assumption of the transferability of data from one to another protein structure, when such data are used in modeling studies.

The maintenance of a unique tertiary structure may at first seem unfavorable, due to the entropic consideration of all the possible conformations available to a polypeptide (Carter, 1998; Dill *et al.*, 1995). This gain in entropy is more than offset by the formation of tertiary contacts (interactions/bonds) formed between remote structural elements of a polypeptide backbone in the globular states assumed by most proteins (Carter, 1998). No

single theory of tertiary structure formation exists presently, although two concepts are suggested (Carter, 1998). Firstly, the polypeptide backbone has a characteristic fold with some degree of imprecision, which relates to the primary structure encoding (Flöckner *et al.*, 1995) and to the mechanism of renaturation (Dill & Chan, 1997). Secondly, stereochemical complementarity (Carter, 1998) is required to stabilise unique native folds. Following on from this complementarity, hydrophobic side chain interactions are usually more significant than intramolecular hydrogen bonds between main chains, when considering the stabilising effects of native tertiary structures (Richards & Lim, 1993). Van der Waals interactions, which consist of long-range attractive and short-range repulsive forces, lead to high packing densities in the former and the avoidance of stereochemical congestion in the latter (Richards & Lim, 1993). The topology of protein tertiary structures would seem to reveal a limited repertoire of states (Carter, 1998) in which α -helices and β -strands can pack together to form tight globular structures. A classification of tertiary structures proposed by Levitt and Chothia (1976), Richardson (1977), and Chou (1995) is illustrated below in table 4.2.

Table 4.2 Table of tertiary structures and secondary structural elements.

Class	Structural Parameters*
α Proteins	$\alpha \geq 40\%$, $\beta \leq 5\%$
β Proteins	$\alpha \leq 5\%$, $\beta \geq 40\%$
$\alpha + \beta$ Proteins	$\alpha \geq 15\%$, $\beta \geq 15\%$; $> 60\%$ of β -strands antiparallel
α/β Proteins	$\alpha \geq 15\%$, $\beta \geq 15\%$; $> 60\%$ of β -strands parallel
ζ Proteins	$\alpha \leq 10\%$, $\beta \leq 10\%$

*Secondary structure definition is that of Kabsch and Sander (1983).

The designations in table 4.2 represent proteins with mostly α -helical structure (α proteins), mostly β -sheet structure (β proteins), proteins with distinct regions of α and β

structure ($\alpha + \beta$ proteins), interleaved segments of α and β structures (α/β proteins) and mostly disordered structures (ζ proteins).

The quaternary structure of proteins describes the association of monomeric protein subunits (whole proteins or polypeptide fragments) into oligomeric assemblies. As with tertiary structures, van der Waals interactions are the main source of complex formation (Carter, 1998), entailing a reduction in rotational and translational entropy, which has to be offset by a large number of contacts (bonds) across the interface between protein subunits. The burial of solvent accessible surface area between subunits is the main driving force of quaternary bond formation, with the proportionality of $\sim 25 \text{ cal/mole/\AA}^2$ of buried surface area observed for complex formation (Chothia, 1974). Questions of symmetry, control and cooperativity all relate to the quaternary structure (Carter, 1998). For example, the approximate dihedral symmetry associated with hemoglobin, a tetramer with two two-fold axes normal to a two-fold principal axis, suppresses the individual polarities of individual subunits to give an approximately isotropic polar surface, which is presented to an aqueous environment (Carter, 1998). Clearly, the ability that a protein has for modulating its polar properties is important for its solubility in different media (e.g. if a protein is transported across a membrane from an aqueous to a less polar environment). In addition, changes in the quaternary structure are responsible for the cooperative binding of oxygen in oxygen transport proteins such as hemoglobin (Shriver *et al.*, 1994). Here, cooperativity is a function of the number of subunits present in a protein (Carter, 1998). As the number of subunits, n , increases the uptake of oxygen also increases (Perutz, 1990). This is made possible if information on the number of binding states is communicated through conformational changes coupled to the ligation state of the protein.

4.1.1. Glycoproteins: Structure of Glycans and their Attachment to Proteins

Glycoproteins are proteins with a complex manifold of oligosaccharides attached through bonds known as glycosidic linkages, with the oligosaccharides known as glycans

(Fukuda, 1994). Six sugars are commonly found in the structures of complex animal glycoproteins: galactose, mannose, N-acetylglucosamine, N-acetylgalactosamine and sialic acid from the D-series, and fructose from the L-series (David, 1997). The linkage types encountered in glycoproteins are either N-glycosidic or O-glycosidic (David, 1997), depending on whether the oligosaccharides are linked through asparagine residues (N-glycans) or serine and threonine residues (O-glycans), respectively. Individually, the asparagine, threonine and serine residues are not sufficient to signal glycosylation (Creighton, 1993). For N-glycosylation to occur a triplet sequence of -Asn-Xxx-Ser-/Thr-/Cys- (Xxx is any residue except proline (Creighton, 1993)) is required as a prerequisite to glycosylation. Such a characteristic sequence is, however, not the only determiner of glycosylation (Creighton, 1993), as Asn sites are not always glycosylated. The situation for O-glycosylation is less clear. As a general rule amino acid sequences are not a 'strong' determiner of glycosylation at serine or threonine residues (Creighton, 1993). Some exceptions do occur, however. For example, in collagen, hydroxy-Lys residues (Hyl) that are glycosylated occur in the characteristic sequence, -Gly-Xxx-Hyl-Xxx-Arg-, where Xxx is any residue (Creighton, 1993). In addition, O-glycosylation normally occurs after a protein has folded; whereas N-glycosylation occurs in a nascent protein during its passage through the endoplasmic reticulum (Creighton, 1993; Rudd & Dwek, 1997).

N-glycans are conveniently divided into three classes: high mannose type, complex type and hybrid type (Fukuda, 1994). Each of these glycans has their reducing terminal N-acetylglucosamine linked to asparagine forming an aspartylglucosamine linkage. The amino group is directly attached to the C-1 carbon atom of the N-acetylglucosamine residue, since the original hydroxyl group attached to the C-1 atom is removed during the formation of this linkage (Fukuda, 1994). All three types of glycans have a common 'core' structure known as the trimannosyl core, $\text{Man}\alpha 1 \rightarrow 6 (\text{Man}\alpha 1 \rightarrow 3) \text{Man}\beta 1 \rightarrow 4 \text{GlcNAc}\beta 1 \rightarrow 4 \text{GlcNAc}\beta 1 \rightarrow \text{Asn}$. Three mannose residues and di-N-acetyldibiose make up the core portion of this structure in which the mannose groups can form branches with $\alpha 1 \rightarrow 6$ or $\alpha 1 \rightarrow 3$ linkages. High mannose-type oligosaccharides may also have an additional two to six mannoses attached to their core structure (Fukuda, 1994). Complex-type oligosaccharides always contain three mannose residues with N-acetylglucosamine and

Galβ1 → 4 GlcNAc attached to the outer two mannose residues, which are followed by sialic acid residues or additional N-acetylglucosamines. Hybrid-type glycans contain more than three mannose residues but also contain N-acetylglucosamines attached to α1 → 3 mannose. Each of these glycan types is illustrated in Figure 4.1.

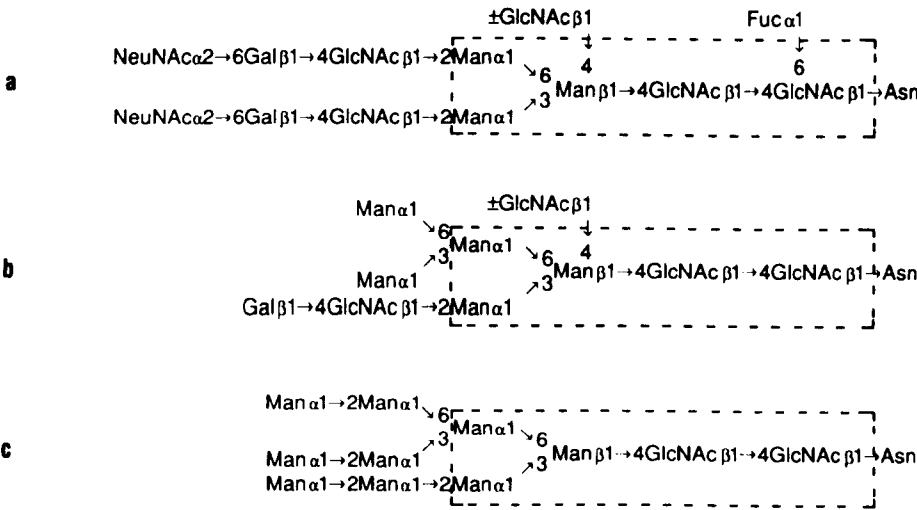


Figure 4.1. The different types of N-glycans. (a) Bi-antennary complex-type. (b) Hybrid-type oligosaccharides. (c) High mannose-type oligosaccharides. (Reproduced from Fukuda, 1994.)

Since O-glycans were originally found in mucus type substances they are also known as mucin-type glycans. Their reducing terminal sugars are attached to hydroxyl groups of polypeptides and, like the N-glycans, they are subdivided into different groups (Fukuda, 1994). The mucin-type carbohydrates with N-acetylgalactosamine linked to serine or threonine, the O-GlcNAc type with N-acetylglucosamine linked to serine or threonine, and proteoglycans with xylose linked to serine or threonine. All three of these glycan types are fairly abundant in the animal kingdom. Other less often encountered O-glycan types are galactose attached to hydroxylysine in collagen, xylosyl-glucose or glucose attached to serine or threonine clotting factors, and fucose attached to plasma glycoproteins. In the plant kingdom arabinose is attached to hydroxyproline, galactose is linked to serine, and in fungi, mannose is linked to serine or threonine. All of the O-glycan

types in the previous discussion can be found in Fukuda (1994). The widespread nature of these glycans is a testament to the great utility of carbohydrates in biological systems.

The role of N- and O-glycans is multifarious, but also not fully understood (Rudd & Dwek, 1997). Complications stem from the fact that heterogeneous populations of glycans appear at sites of glycosylation making it difficult to differentiate the influence of individual glycans on a protein's structural and physical properties. X-ray crystallography has difficulties in defining the sugar conformations due to the mobility and heterogeneity of the oligosaccharide residues. However, NMR does not suffer from this problem, with the proviso that only the major glycoforms are accessible to study (Rudd & Dwek, 1997). A database of oligosaccharide structural units, attached to the protein, can be used to model the conformational space used by the glycans. This is justified by the fact that oligosaccharide conformations are generated from their secondary structure, unlike proteins where tertiary and quaternary interactions play a significant role (Rudd & Dwek, 1997).

Some of the properties that glycosylation confers on a protein are now discussed. One of the most interesting properties of glycoproteins is their kinetics of folding/unfolding and their thermal stabilities (Rudd & Dwek, 1997; Arnold & Ulbrich-Hormann, 1997). Kinetic stabilities (ΔG^\ddagger) for ribonuclease A and B (Arnold & Ulbrich-Hormann, 1997), calculated from unfolding constants, show that the glycosylated B form is kinetically more stable, resulting from enthalpic effects, whereas entropic effects tend to counteract stabilisation. Conformational studies (Rudd & Dwek, 1997) of oligosaccharides indicate that glycosylation can alter the conformational profile of a folding polypeptide, which allows alternative folding pathways to be followed, producing structural nucleation sites that are not found in an unglycosylated protein. Catalytic effects are also affected by glycosylation, with hydrogen tunneling and enthalpy of activation barriers being increased with a corresponding increase in the extent of glycosylation (Joao & Dwek, 1993; Rudd & Dwek, 1997). In addition, the inhibition of proteolytic enzymes is achieved if sites that are susceptible to attack are glycosylated (Rudd & Dwek, 1997). The solubility of highly concentrated glycoproteins in the lumen of the endoplasmic reticulum is enhanced by the hydrophilic properties of the sugars, which may prevent aggregation of

the newly formed protein (Rudd & Dwek, 1997).

This section on glycan structure and function is necessarily only a brief introduction, due in part to the lack of significant glycan bands seen in glycoprotein ROA spectra. It was found, during the course of this project, that only a very few carbohydrate features appear in the ROA spectra of glycoproteins and was not in any obvious sense a function of the carbohydrate content of the proteins studied. Some glycoproteins with a low carbohydrate content have one or two bands attributed to sugars, whereas other glycoproteins with a large carbohydrate content have fewer sugar bands than one might expect. As a result of this, greater emphasis, therefore, is placed on ROA assignments of the polypeptide conformation in protein structure.

It was not possible to conduct studies on appropriate model carbohydrates as they are expensive. Typically, a 20 mg sample is required to obtain a protein spectrum, however, carbohydrates usually require higher concentrations (e.g. glucose requires ~ 30 to 40 mg to acquire a ROA spectrum). The cost of model carbohydrates (e.g. high mannose-type oligosaccharides) is prohibitive, with a 20 mg sample costing ~ £1000. Also, the use of endoglycosides (Thotakura & Bahl, 1987) in deglycosylation studies, to produce samples of deglycosylated protein and glycan for ROA studies, was also considered prohibitive. For example, endoglycosides costing ~ £100 will deglycosylate ~ 2mg of glycoprotein (Thotakura & Bahl, 1987; Seikagaku, AMS Biotechnology). To obtain a suitable amount of glycan for a ROA study would require large amounts of glycoprotein (~ 200 to 500 mg), depending on the initial amounts of glycan(s) present in a glycoprotein sample.

4.2. Normal Modes of Proteins

The collective simple harmonic motion of the constituent atoms about their equilibrium position in a molecule is referred to as the characteristic normal modes of vibration (Wilson *et al.*, 1955), with the normal modes of water and carbon dioxide being the classic examples encountered by all chemistry students. However, unlike these simple

molecules with their relatively few vibrational modes (three and four modes respectively), protein molecules will have a vast number of normal modes. As each atom in a molecule has three degrees of freedom (x, y, z) a molecule with N atoms will have 3N degrees of freedom, with appropriate rotational and translational constraints deducted from this figure to give the 3N - 6 normal modes of vibration (Wilson *et al.*, 1955). Clearly, as the size of a protein increases the number of normal modes of vibration will increase, resulting in a challenging problem for computational chemists to accurately calculate the frequencies and intensities of spectral transitions associated with the normal modes of a protein. The earliest relevant model for proteins, developed by Miyazawa *et al.* (1958), involved the use of N-methylacetamide (NMA) as a simple model of the peptide group motion, with a total of nine bands predicted in its Raman spectrum. Recent and more subtle experiments and calculations have refined this model (Asher *et al.*, 1998; Chen *et al.*, 1995). However, the size of this molecule naturally places a limit on the range of applicability in analysing complex protein structures.

The current state-of-the-art in normal mode analysis calculations can be attributed to the work of Krimm and his coworkers (Krimm & Bandekar, 1986; Krimm & Reisdorf, 1994). Over the past twenty to thirty years this work has involved the development of empirically determined, conformation-dependent peptide and polypeptide force fields, as well as ab-initio and molecular mechanics force fields (Krimm & Reisdorf, 1994). For example, in a molecular mechanics force field the potential energy, V, is given by a quadratic bonded term, V_q (the harmonic approximation is assumed) and a non-quadratic term, V_{nq} , yielding

$$V = V_q + V_{nq} \quad (4.1.1)$$

with

$$V_q = \frac{1}{2} \sum_i F_{ii} (q_i - q_{i0})^2 + \sum_{i < j} F_{ij} (q_i - q_{i0})(q_j - q_{j0}) \quad (4.1.2)$$

and

$$V_{\text{nt}} = \sum V_{\text{tor}} + \sum V_{q,\text{tor}} + \sum V_{\text{tor},\text{tor}} + \sum V_{\text{nb}} \quad (4.1.3)$$

The q_i in equation 4.1.2 are internal coordinates with q_{i0} as reference values. The form of a typical torsion term is

$$V_{\text{tor}} = \frac{1}{2} V_N (1 \pm \cos N\phi) \quad (4.2.1)$$

where N is the periodicity and ϕ is the torsion angle. For empirical force fields similar equations can be derived (Krimm & Reisdorf, 1994). In the past the parameterisation of these force fields was achieved by obtaining agreement with structures and energies (Krimm & Reisdorf, 1994), with little attention being paid to spectral vibrational data such as band frequencies. Recently, spectroscopically determined force fields (SDFFs) have been developed with greater emphasis placed on incorporating frequency agreements into molecular mechanics force fields at the start of the parameter optimisation process. This means that the errors usually encountered with non-SDFFs (e.g. frequency errors as great as 50 cm^{-1}) are removed with the agreement now $\sim 2 \text{ cm}^{-1}$ for SDFFs and ~ 5 to 10 cm^{-1} for empirical force fields (Krimm & Reisdorf, 1994).

From the previous discussion it is tempting to think that the problems associated with force field calculations are all but dealt with; however, a few problems remain. The only allowed variations in peptide geometry, for example, are $\phi(\text{N-C}^\alpha)$ and $\psi(\text{C}^\alpha\text{-C})$ (Pauling & Corey, 1951a). Also, although variations from planar geometry are known to occur in proteins, the peptide backbone is fixed in a planar position (Krimm & Reisdorf, 1994). The role of intercalated water in protein secondary structures is also of great concern. Recent studies on NMA suggest that the bending modes of water may couple vibrationally to the amide I mode of the protein backbone through hydrogen bonds (Chen *et al.*, 1995), which could be inextricably linked with the dynamics of proteins. Also, throughout this discussion it was assumed that protein side chains are approximated as

point masses (Krimm & Bandekar, 1986; Krimm & Reisdorf, 1994; Krimm, 1987).

In practical terms the gleaning of structural information from ROA protein spectra is achieved by correlating structural evidence from independent sources such as CD, Raman, NMR and X-ray crystallography with model peptides (Birke *et al.*, 1992; Carey, 1982; Evans, 1995; Carter, 1998). For example, the homopolypeptide poly(L-lysine), which is one of the most extensively studied peptides spectroscopically, is seen to have a number of differences in its CD and Raman spectra under differing states of pH and temperature (Birke *et al.*, 1992; Carey, 1982). Intuitively, the differences are viewed as representing rearrangements in the structure of the polypeptide. Tentative band assignments are made and then compared with other spectra with known structural elements. Proceeding in this manner a database of spectral assignments is built-up.

Chapter 5

Immunoglobulin G and Other Proteins with Mostly β -sheet Structure

One of the objectives of this project is the search for ROA band patterns characteristic of similar structural motifs in proteins. If a protein sample is difficult to crystallise, the sample may be more amenable to ROA studies, which would allow for structural studies against samples with known structures. By looking at an ROA spectrum structural motifs could be ‘picked’ out to provide clues on the fold of a protein. In this chapter spectra of immunoglobulin G (IgG), avidin and pepsin are presented, with emphasis given to IgG due to its importance in immunology.

5.1. Immunoglobulin G (IgG)

Immunoglobulin G is a glycoprotein with ~ 1400 amino acid residues, a mass of ~ 150 kDa, and a carbohydrate content of ~ 3 to 5% (Rademacher *et al.*, 1996). It is one of a number of molecules produced during an immune response, with the others being immunoglobulin A, immunoglobulin D, immunoglobulin E and immunoglobulin M (Roitt, 1997). IgG is regarded as being the most important immunoglobulin due to its relative abundance and ease of diffusion into body spaces, where it neutralises bacterial toxins and enhances phagocytosis responses upon binding to microorganisms (Roitt, 1997). As a testament to the importance of IgG a glance at the protein databank (PDB) (Bernstein, 1977) reveals that of the some 6000 or so entries ~ 120 (at the last count) are of diffraction studies on IgG and its fragments. The structure of IgG is rather elegant. Schematically its structure is given in Figure 5.1 and is seen to have a Y-shape with C_2 symmetry (Roitt, 1997).

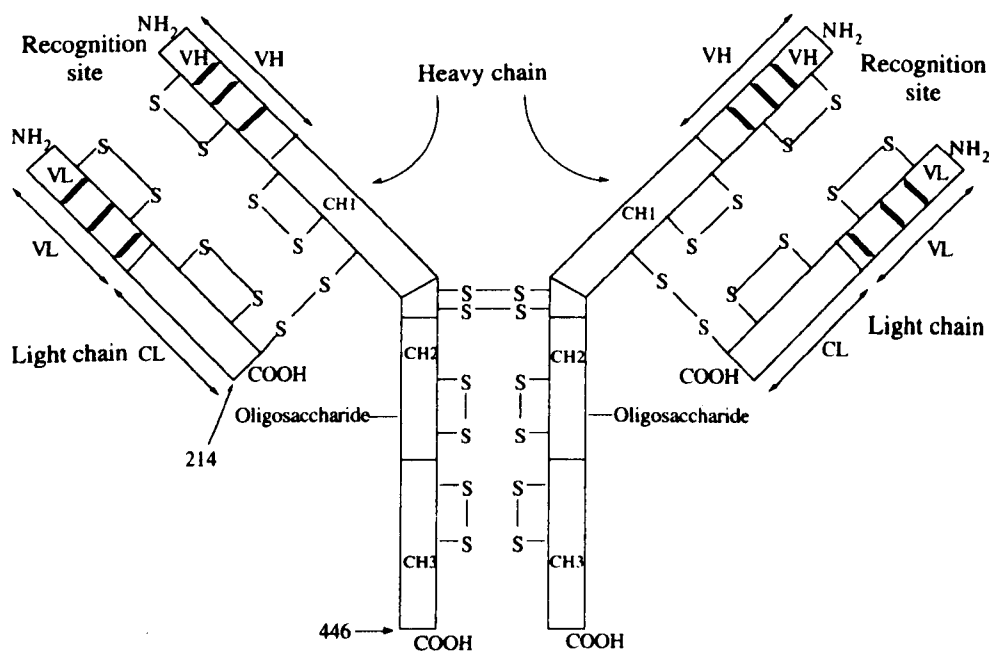


Figure 5.1 Schematic representation of IgG with its Y-shape C_2 symmetry. (Reproduced from David, 1997.)

It can be seen that IgG is divisible into a number of structural components (Roitt, 1997; Marquart & Deisenhofer, 1982). The two arms are known as F_{ab} fragments, and are joined to what is called the F_c stem. The polypeptide structure of IgG consists of a short chain known as the light chain (L) and a long chain known as the heavy chain (H), with both light chains identical to each other and similarly for the heavy chains. There are six domains, with the designations: V_H (variable heavy chain domain), V_L (variable light chain domain), C_{H1} (1st constant heavy chain domain), C_L (constant light chain domain) and similarly for C_{H2} and C_{H3} . In addition, from the MOLSCRIPT diagram (Figures 5.2 & 5.3) (Kraulis, 1991), it can be seen that IgG is a β -sheet protein with an average anti-parallel β -sheet content of ~ 45 to 50% (Saul & Poljack, 1991), with the bulk of the remaining structure consisting largely of β -turn.

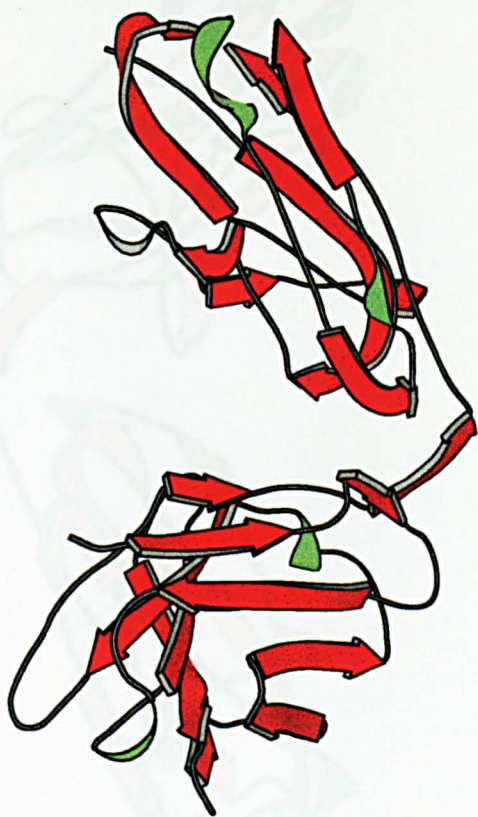


Figure 5.2 MOLSCRIPT diagram of F_{ab} fragment 8fab, which shows the heavy chain, H.

Previous Raman studies on IgG (Painter & Koenig, 1975; Pérez et al., 1976) were used as reference spectra and were found to be in agreement with the parent Raman spectrum of the current work. Table 5.1 is given below to compare these results.

Table 5.1. Table of Raman assignments of IgG.

Frequency (cm ⁻¹)/H ₂ O	Frequency (cm ⁻¹)/D ₂ O	Frequency (cm ⁻¹)/H ₂ O	Frequency (cm ⁻¹)/D ₂ O	Frequency (cm ⁻¹)/H ₂ O	Assignment
621	621	621	621	623	Phe
643	644	643	643	644	ν C-SH-Tyr
737	738	737	737	739	Trp
830	830	830	830	832	Tyr
851	851	851	851	857	Tyr
878	881	878	877	878	Trp
931	931	931	931	933	ν C-O
958	962	958	960	959	ν C-O
	990		991		ν C-O
1000	1004	1000	1004	1005	Phe
1028	1030	1028	1032	1012	Phe + Trp
1057	1062	1057	1064	1034	Phe
1124	1129	1124	1132		ν C-N
1130	1160 (sh)	1130	1132	1126	ν C-N
1174	1171	1174	1174	1176	Tyr + Phe
1200	1205	1200	1208	1210	Tyr + Phe
1230	1230 (sh)	1230	1230	1237	Amide III
				1239	Amide III
				1267	
1304 (sh)		1304	1304		
1315	1321 (sh)	1315	1317	1325	C α -NH

Figure 5.3 MOLSCRIPT diagram of F_{ab} fragment 8fab, which shows the light chain, L.

Previous Raman studies on IgG (Painter & Koenig, 1975; Pézolet *et al.*, 1976) were used as reference spectra and were found to be in agreement with the parent Raman spectrum of the current work. Table 5.1 is given below to compare these results.

Table 5.1. Table of Raman assignments of IgG.

Frequency (cm ⁻¹) ^a H ₂ O	Frequency (cm ⁻¹) ^a D ₂ O	Frequency (cm ⁻¹) ^b H ₂ O	Frequency (cm ⁻¹) ^b D ₂ O	Frequency (cm ⁻¹) ^c H ₂ O	Assignment [*]
621	621	622	621	623	Phe
643	644	646	643	644	$\nu(\text{C-S})$ + Tyr
757	758	760	758	759	Trp
830	830	830	830	832	Tyr
853	851	856	856	857	Tyr
878	881	879	877	878	Trp
931	931	-	-	933	$\nu(\text{C-C})$
958	962 (sh)	955	960	959	$\nu(\text{C-C})$
-	990	991	991	-	$\nu(\text{C-C})$
1000	1004	1004	1004	1005	Phe
-	-	-	-	1013	Phe + Trp
1028	1030	1031	1032	1034	Phe
1077	1082	1078	1084	-	$\nu(\text{C-N})$
1124	1129	1123	1122	1126	$\nu(\text{C-N})$
1155	1160 (sh)	1152	-	1160	-
1174	1171	1174	-	1176	Tyr + Phe
1207	1205	1208	-	1210	Tyr + Phe
1238	1230 (sh)	1240	-	1237	Amide III
-	-	-	-	1259	Amide III
-	-	-	-	1267	-
1288 (sh)	-	-	-	-	-
1321	1321 (sh)	1318	1317	1325	CH ₂ def.
1340	1338	1336	1336	1340	Trp + C $_{\alpha}$ -H

1357 (sh)	1363	1359	1359	1360	Trp + C α -H
1399	1385	1400	1406	1400	$\nu_{\text{sym}}(\text{COO}^-)$
1448	1450	1449	1451	1450	CH ₂ def.
1511	-	-	-	-	Trp
-	1553	1557	1554	1554	Trp
1581 (sh)	1587 (sh)	-	1584	-	Trp
-	-	-	-	1606	Trp, Phe
1617	1611	-	1613	1606	Trp, Tyr, Phe
1670	1664	-	1667	1673	Amide I, I'

Raman assignments from current Raman and ROA study of IgG^a; Raman assignments from Painter and Koenig^b (1975) and from Pézolet^c et al. (1976), Sh: shoulder. ^aAll of these assignments are from (Painter & Koenig, 1975; Pézolet et al., 1976).

The Raman and ROA spectra of IgG are shown in Figure 5.4. As table 5.1 indicates, the Raman spectrum of IgG is dominated by aromatic side chain bands (see chapter 1). The ROA spectrum, however, is dominated by only those vibrational modes that are most directly involved in the polypeptide backbone. As is common with most protein ROA spectra the region from $\sim 650 \text{ cm}^{-1}$ to 850 cm^{-1} has little discernible band structure. The positive band at $\sim 950 \text{ cm}^{-1}$ is normally assigned as α -helix (Wilson, 1996), which seems at odds with X-ray crystallography (Saul & Poljack, 1991), and Raman studies (Painter & Koenig, 1975; Pézolet et al., 1976). Early Raman studies suggest that the bands in the 930 cm^{-1} to 990 cm^{-1} wavenumber region are from C-C stretching modes from side chain groups and not from the backbone (Painter & Koenig, 1975; Pézolet et al., 1976). Furthermore, disordered proteins are also known to have Raman bands in this region (Frushour & Koenig, 1974). Unfortunately, there is no easy way to test these proposals. If the 950 cm^{-1} band is representative of disordered structure it may be suggested that deuteration studies would reveal this, as such a band would be expected to disappear upon dissolution in D₂O (Wilson, 1996; Tu, 1986). Here, the D₂O ROA spectrum is unaffected below $\sim 950 \text{ cm}^{-1}$.

Next the parent Raman band at $\sim 1009 \text{ cm}^{-1}$ is assigned as either a phe or a trp ring breathing mode (Painter & Koenig, 1975; Pézolet et al., 1976; Tu, 1986) and is quite prevalent in ROA spectra, but has yet to be assigned any structural significance (see

discussion on avidin). The ROA couplet, negative at $\sim 1080\text{ cm}^{-1}$ and positive at $\sim 1122\text{ cm}^{-1}$ is associated with β -sheet proteins with Raman studies assigning both of these bands as C_{α} -C and C_{α} -N stretches (Painter & Koenig, 1975; Pézolet *et al.*, 1976). However, the ROA spectrum of avidin does not show this band (avidin has a comparable β -sheet content to IgG). ROA studies suggest that this region is also associated with α -helix (Wilson, 1996), but the α -helix content of IgG is rather low (~ 3 to 5%) (Fagnano & Fini, 1993). Stretches of irregular α -helix, which are not uncommon in IgG (Chothia & Lesk, 1987), may be contributing to this band.

In previous ROA studies, features in the extended amide III region from ~ 1220 to 1360 cm^{-1} have been a source of much discussion (Wilson, 1996). A prominent negative band located between $\sim 1240\text{ cm}^{-1}$ and 1245 cm^{-1} in the ROA spectrum (corresponding Raman band $\sim 1238\text{ cm}^{-1}$) is now assigned as individual β -strands, whereas previously it was tentatively thought to be PPII loop structure (Wilson, 1996), a type of helical structure contiguous with the β -sheet region of Ramachandran space (Ramachandran *et al.*, 1963; Creamer, 1998). When deuterated the 1240 cm^{-1} ROA band disappears to some extent, suggesting that the structural element associated with this band is solvent exposed, unlike β -sheet structure which is thought to be solvent inaccessible (Jefferey & Saenger, 1994). However, the decrease in this band in the D_2O ROA and Raman spectrum is easily explained by looking at the symmetry of the Raman band at $\sim 1238\text{ cm}^{-1}$. In a Raman study by Frushour and Koenig (1974) it was indicated that partially disordered β -sheets produce an asymmetric band (asymmetric on the high wavenumber side of the β -sheet band). This asymmetry is clearly seen in the Raman spectrum, shown in Figure 5.4, and indicates irregularities in the β -sheet structure. Also, there is a $\sim 1210\text{ cm}^{-1}$ to 1220 cm^{-1} negative band, which may also be associated with β -structures. IgG has a barrel architecture with seven anti-parallel strands (Roitt, 1997). In a subsequent chapter the ROA spectrum of aldolase, which has a different barrel structure, is presented and is seen to have a similar feature, but with a different barrel structure. Small negative bands appearing as shoulder at $\sim 1257\text{ cm}^{-1}$ and 1271 cm^{-1} also consistently appear in the ROA spectrum of IgG. Raman studies from the mid 1970s assigned bands in this region as disordered structure (Painter & Koenig, 1975; Pézolet *et al.*, 1976). However, more recent studies by Krimm & Bandekar (1980, 1986) have suggested that Raman bands

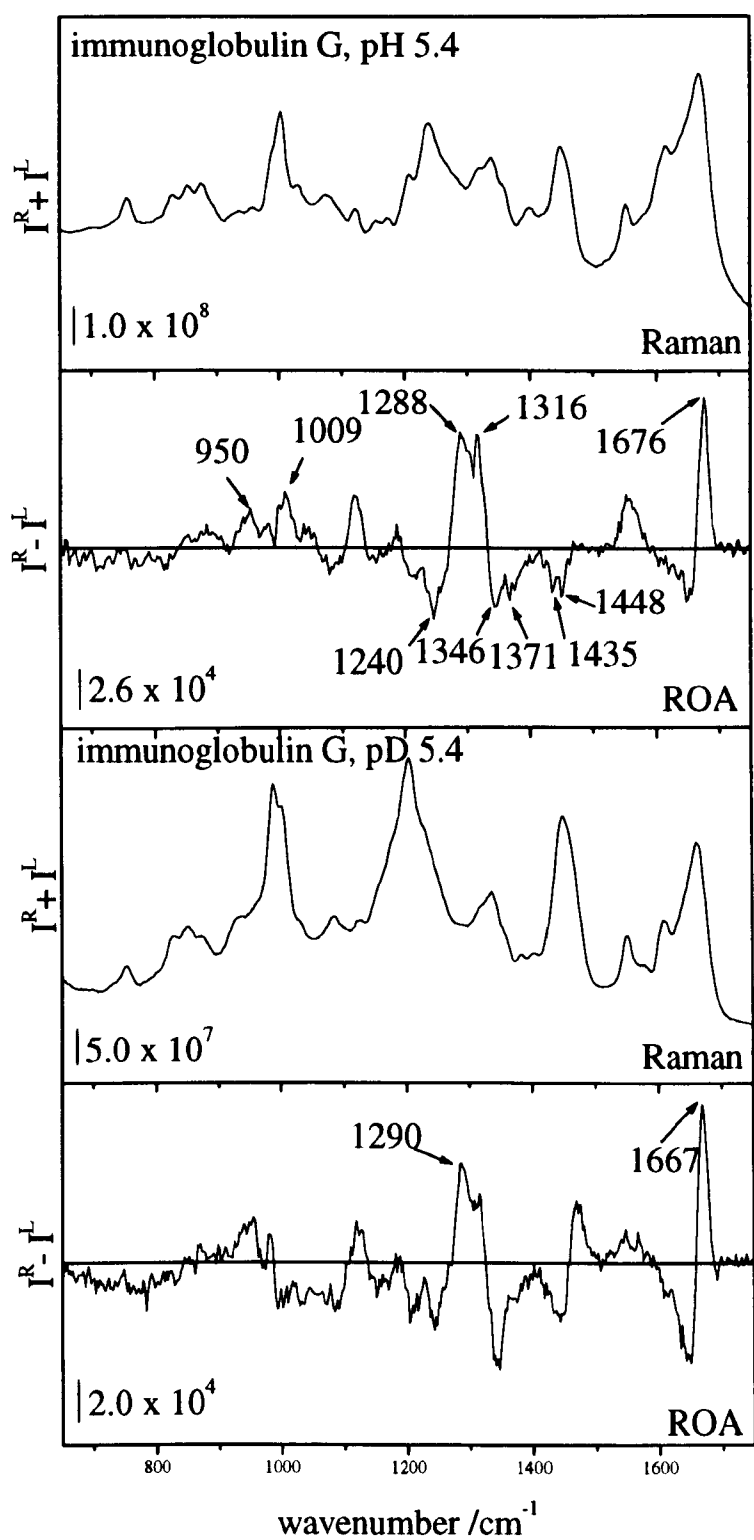


Figure 5.4 Raman and ROA spectra of Human IgG in H₂O (top) and IgG in D₂O (bottom).

in the general region from $\sim 1260\text{ cm}^{-1}$ to 1290 cm^{-1} are of different β -turn structures, especially type I β -turn, as this is the most common, along with type III (Perczel & Jákli, 1995). The strong positive band at $\sim 1288\text{ cm}^{-1}$ is, in fact, in a region assigned as type I and III turn structure by Krimm & Bandekar (1980, 1986). One other possible interpretation of this band, which ties in with the proposal of disordered structure, is that individual amino acid residues, possibly in long loops connecting the β -sheet domains (see MOLSCRIPT diagrams), are clustering in a region of Ramachandran space with local order that of α -helix (Perczel & Jákli, 1995). It is readily seen in Figure 5.4, from the D_2O ROA spectrum, that the 1288 cm^{-1} band disappears to some extent leaving behind an asymmetric band at $\sim 1290\text{ cm}^{-1}$, which originates from C-H deformations (Diem, 1993). Similarly, a second positive band at $\sim 1316\text{ cm}^{-1}$ disappears to an even larger extent after deuteration. Recent ROA studies on disordered homopolypeptides such as poly(L-lysine) and poly(L-glutamic acid) suggest that this band is acting as a marker of disordered structure, possibly originating in PPII loop structure (Smyth *et al.*, to be published; Wilson *et al.*, 1996b). Two negative bands at ~ 1346 and 1371 cm^{-1} are next seen in the ROA spectrum and are usually associated with β -turn structure (Wilson, 1996), with the 1371 cm^{-1} band partially removed from the ROA spectrum after deuteration. Interestingly, the 1346 cm^{-1} band is enhanced quite considerably in D_2O and is often seen when samples with turn structure are deuterated (Wilson, 1996). Conventional Raman studies have assigned this band as a $\text{C}_\alpha\text{-H}$ deformation (Overman & Thomas, 1998) and it is not inconceivable that this band is masked in H_2O samples by significant contributions from N-H deformations (Wilson, 1996). Next, two bands at ~ 1435 and 1448 cm^{-1} are seen in the ROA spectrum, and are assigned as a tryptophan band and CH_2 deformation, respectively (from Raman studies), and demonstrate that although ROA is most sensitive to backbone conformation details, significant ROA can sometimes be generated by side chain vibrations. The amide I region of the IgG spectrum is dominated by C=O stretches ($\sim 85\%$) (Krimm & Bandekar, 1986), and shows a ROA couplet, negative at $\sim 1653\text{ cm}^{-1}$ and positive at $\sim 1676\text{ cm}^{-1}$ centred at $\sim 1662\text{ cm}^{-1}$. The positive peak correlates with the usual Raman assignment of β -sheet at $\sim 1674\text{ cm}^{-1}$ (Painter & Koenig, 1975; Pézolet *et al.*, 1976). Also, in D_2O , this band shifts to $\sim 1667\text{ cm}^{-1}$ and corresponds exactly with the amide I' assignment of β -sheet structure (Painter &

Koenig, 1975; Pézolet *et al.*, 1976). In observing the amide I' band, it can be seen that no intensity is lost, which indicates the significant C=O contribution in this band.

In chapter one it was mentioned that methods exist for the determination of protein secondary structure content from Raman spectra. Pézolet *et al.* (1976) estimated the β -sheet content of IgG to be $\sim 37\% \pm 4\%$, from the parent Raman spectrum. However, the relative accuracy of this method is no better than $\pm 10\%$, and depends on the assumption that the 1450 cm^{-1} band, which is used in this method, is independent of protein secondary structure vibrations.

5.2. Avidin

The whites of hen eggs are rich in protein and glycoprotein fractions, with one such component being avidin (Livnah *et al.*, 1993). It is a tetramer consisting of subunits with masses of $\sim 15.7\text{ kDa}$ and is also a glycoprotein with a carbohydrate content of $\sim 12\%$. The biological role of avidin is unclear. However, in an early study by Green (1975), it was suggested that an antibacterial function is likely, as bacterial growth is inhibited in hen eggs when D-biotin (vitamin H) is sequestered by avidin. The avidin-biotin complex forms one of the strongest complexes known to biochemists (dissociation constant $K_D = 6 \times 10^{-16}\text{ M}$) (Torreggiani & Fini, 1998) and, consequently, is of great interest both theoretically and experimentally. The avidin-biotin complex has been used as a model of the IgG-antigen interaction and is useful as a biochemical tool in the affinity isolation of proteins containing biotin derivatives (Honzatko & Williams, 1982).

The X-ray crystal structure of avidin reveals an anti-parallel, eight-stranded β -barrel with the remainder of the structure consisting of β -turns and a small fraction ($\sim 2.5\%$) of 3_{10} -helix (data from pdb, 1avd). The percentage of β -sheet is estimated to be $\sim 49\%$ (Pugliese *et al.*, 1984).

As with IgG, the low wavenumber region of the ROA spectrum of avidin is devoid of any bands. Furthermore, unlike IgG, and in keeping with the crystal structure, no 'anomalous' α -helix bands appear. The first prominent feature in the ROA spectrum

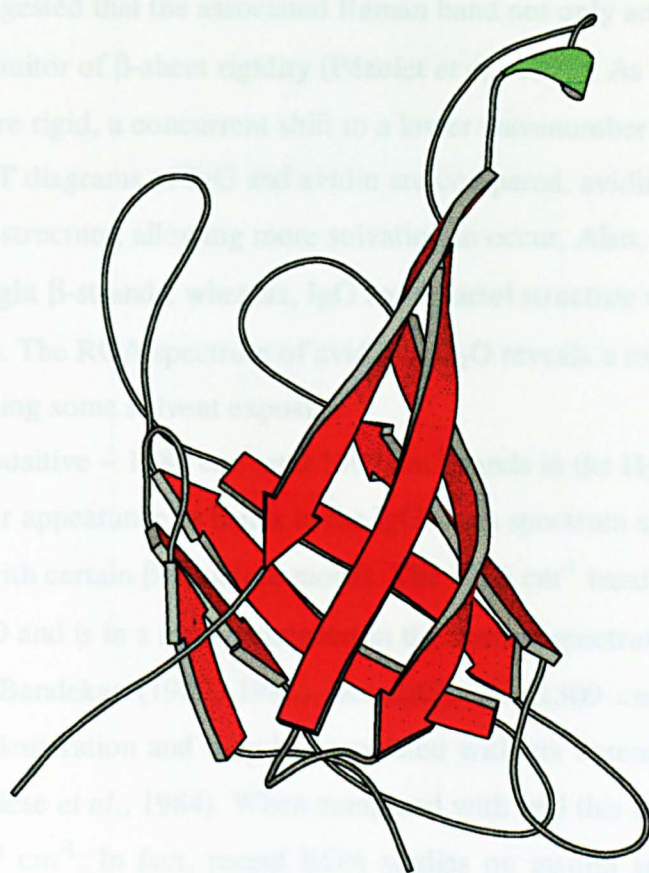


Figure 5.5 MOLSCRIPT diagram of avidin, 1avd. (Wilson, 1980, 1985).

is the positive tryptophan band at $\sim 1010\text{ cm}^{-1}$, which has a similar appearance to the same band in IgG. Raman studies of proteins often emphasise the importance of tryptophan as a marker band of its local environment, that is, hydrophilic *versus* hydrophobic environments (Carey, 1982). The amide III β -sheet ROA band of avidin, however, differs from IgG in some respects. The centre of this negative feature is less well resolved and is shifted to a lower wavenumber at $\sim 1238\text{ cm}^{-1}$ (IgG $\sim 1245\text{ cm}^{-1}$). It has been suggested that the associated Raman band not only acts as a marker of β -sheet, but also a monitor of β -sheet rigidity (Pézolet *et al.*, 1976). As the β -sheet structure becomes more rigid, a concurrent shift to a lower wavenumber is also observed. If the MOLSCRIPT diagrams of IgG and avidin are compared, avidin would seem to have a more 'open' structure, allowing more solvation to occur. Also, the barrel structure of avidin has eight β -strands, whereas, IgG has a barrel structure with seven β -strands (Roitt, 1997). The ROA spectrum of avidin in D_2O reveals a reduction in the 1238 cm^{-1} band, indicating some solvent exposure.

The positive $\sim 1286\text{ cm}^{-1}$ and 1309 cm^{-1} bands in the H_2O ROA spectrum both have a similar appearance to bands in the IgG ROA spectrum and may be characteristic of proteins with certain β -structure motifs. The 1286 cm^{-1} band is diminished to some extent in D_2O and is in a region assigned in the Raman spectrum as β -turn structure by Krimm and Bandekar (1980, 1986) (see IgG). The 1309 cm^{-1} is reduced to a greater extent after deuteration and may be associated with six extended loops that connect the strands (Pegliese *et al.*, 1984). When compared with IgG this band appears to be shifted down by $\sim 7\text{ cm}^{-1}$. In fact, recent ROA studies on insulin suggest that a similar band located at $\sim 1308\text{ cm}^{-1}$ originates in α -helix (Wilson, 1996). Since its crystal structure reveals no α -helix, it seems unlikely that the same assignment could be made for avidin, although some (~ 2.5 to 5%) 3_{10} -helix is present (data from pdb: 1avd and 1avi). Rather, it is proposed that, in this example, the 1309 cm^{-1} band originates from exposed loop structure with individual residues having propensities for (ϕ, ψ) angles in the α -helix region of Ramachandran space. An alternative proposal is that β -turn signals are common in this region of the ROA spectrum, as they have been seen in the Raman spectra of model polypeptides (Krimm & Bandekar, 1980, 1986).

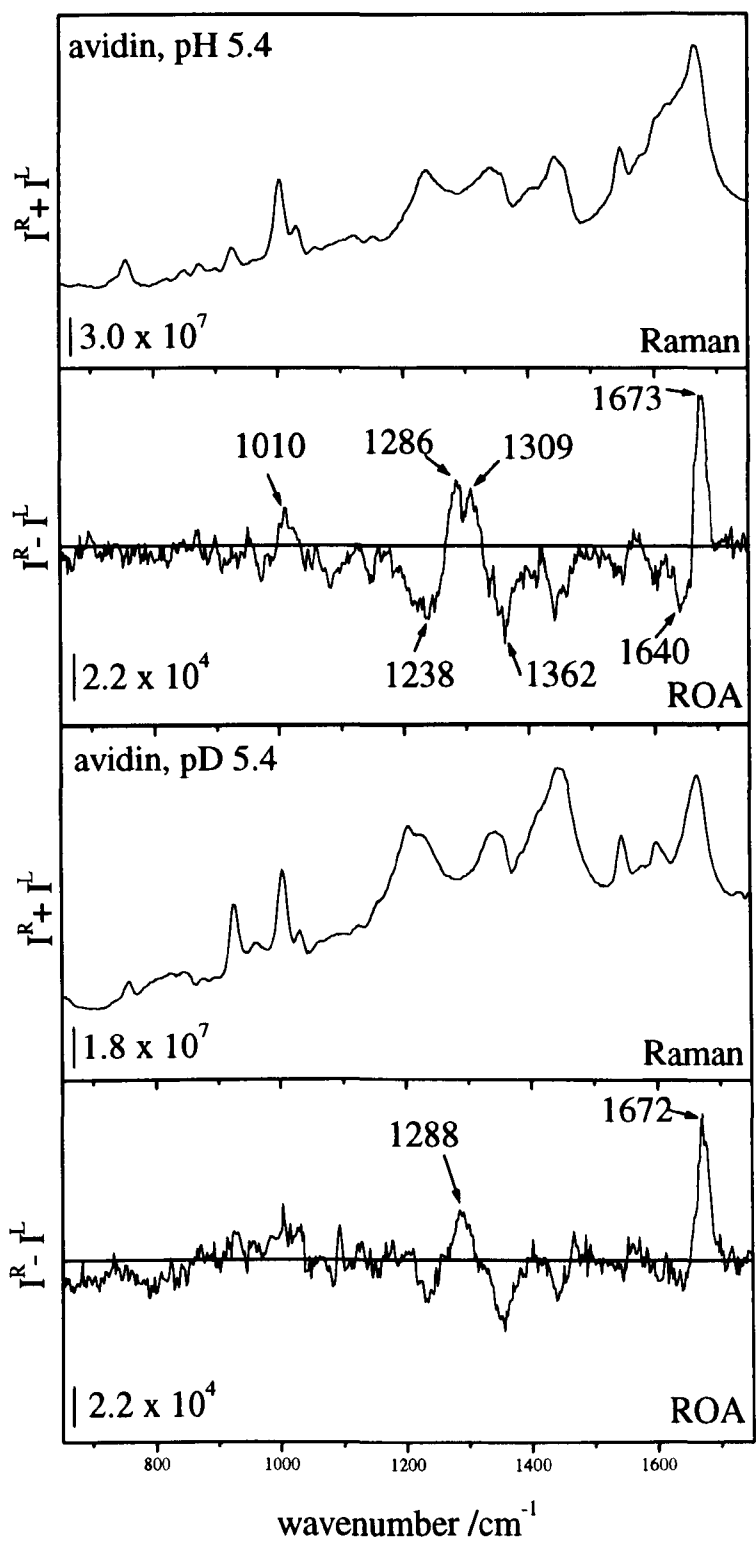


Figure 5.6 Raman and ROA spectra of avidin in H_2O (top) and in D_2O (bottom).

The strong negative peak at $\sim 1362\text{ cm}^{-1}$ in the ROA spectrum is typically assigned as originating from β -turn (Wilson, 1996). However, in Raman studies the 1360 cm^{-1} band is used as a marker of buried tryptophan residues (Painter & Koenig, 1975; Pézolet *et al.*, 1976; Carey, 1982). If the band is strong and sharp, it is suggested that the tryptophan residues are buried and, if the band is weak or absent, the residues are solvent exposed. In the Raman spectrum of avidin this band is weak and suggests solvent exposure of the ~ 16 tryptophan residues (four residues per monomer, pdb: 1avd). However, in the Raman spectrum of IgG, the same observation is also made, with the complication that of the ~ 22 tryptophan residues in IgG, twenty of the residues are buried (Pézolet *et al.*, 1976). In the ROA spectrum of β -lactoglobulin a similar negative ROA band appears at $\sim 1360\text{ cm}^{-1}$ (Blanch *et al.*, 1999), however, β -lactoglobulin has only two tryptophan residues (Blanch *et al.*, 1999). Therefore, it is believed that the negative 1360 cm^{-1} ROA band can be assigned unambiguously as indicating β -turn structure.

In the amide I of the ROA spectrum a negative/positive couplet centred at $\sim 1662\text{ cm}^{-1}$ indicates a high β -sheet composition. The strong positive band at $\sim 1673\text{ cm}^{-1}$ corresponds well with the same band in IgG. However, the negative band is shifted down by 10 cm^{-1} to $\sim 1640\text{ cm}^{-1}$ and may be reflecting subtle differences in the β -sheet composition of avidin *versus* IgG.

5.3. Pepsin

A major component of the digestive enzymes found in the stomach is pepsin, an acid protease, which has evolved to function under the low pH ($\sim \text{pH } 2$) conditions of the stomach (Perutz, 1992), with such conditions denaturing many other proteins. Also, crystallographically, it was one of the first proteins to be studied, but did not have its crystal structure fully analysed until 1977 by Hsu *et al.* (1977) and Subramanian *et al.* (1977). It is hardly surprising that it took so long to elucidate the structure of pepsin. Even at a pH of ~ 1 pepsin carries a net negative charge and has an isoelectric point below ~ 1 (Perutz, 1992). Typical proteins have pK_{a} s of ~ 3.6 to 4.5 and would be

expected to be protonated and fully discharged under similar conditions (Perutz, 1992). The secondary structure of pepsin is dominated by β -sheet structure ($\sim 42\%$), illustrated in Figure 5.7, with \sim six short stretches of α -helix ($\sim 10\%$) and \sim five 3_{10} -helices ($\sim 5\%$) (Adab-Zapatero *et al.*, 1990). The remainder consists largely of turn structure. Two β -sheet lobes separate the protein into two domains, with residues 1-175 forming the N-terminal lobe and residues 176-326 constituting the C-terminal lobe (Adab-Zapatero *et al.*, 1990).

A cleft between these lobes contains a catalytic site, consisting of two aspartate residues on either side of the cleft and a single water molecule bound between both residues (Perutz, 1992). The entrance to the cleft is shielded by two long mobile β -strands (they control access and binding of the substrate) that protrude from either lobe. When a protein is denatured in the acid environment of the stomach, its unfolded polypeptide backbone is inserted into the active cleft where cleavage of the polypeptide occurs.

A striking feature of pepsin's tertiary structure is its similarity to HIV protease (Perutz, 1992), with the main difference being that both lobes in HIV protease are identical, resulting in a pepsin-like fold.

In the H_2O ROA spectrum of pepsin, shown in Figure 5.8, three positive peaks dominate the backbone region. The band at $\sim 947\text{ cm}^{-1}$ is typical for proteins with α -helix. The lower band at $\sim 900\text{ cm}^{-1}$ is also in the α -helix region. The third band in this region is resolved into three peak at $\sim 981\text{ cm}^{-1}$, $\sim 990\text{ cm}^{-1}$ and $\sim 1002\text{ cm}^{-1}$. The 1002 cm^{-1} band originates either in phenylalanine or tryptophan (Painter & Koenig, 1975; Pézolet *et al.*, 1976), with both of the lower bands assigned in the Raman spectrum as originating in $\text{C}_\alpha\text{-N}$ stretching modes from β -sheet (Tu, 1986).

In the extended amide III region of the ROA spectrum two negative peaks at $\sim 1222\text{ cm}^{-1}$ and 1245 cm^{-1} indicate the presence of β -sheet. The origins of the 1222 cm^{-1} band is still unknown, but appears frequently in proteins with a high β -sheet content. It may be a β -sheet variant associated with hydrated β -strands. Krimm has suggested that bands in this region, in Raman spectra, originate in $\text{C}_\alpha\text{-N}$ stretches from type I β -turn (Krimm & Bandekar, 1980; Krimm & Bandekar, 1986). The 1245 cm^{-1} ROA band (centre) is rather broad and spreads over a $\sim 20\text{ cm}^{-1}$ range ($\sim 1236\text{ cm}^{-1}$ to $\sim 1256\text{ cm}^{-1}$). Many of the strands in pepsin are rather short, with two to four residues per β -strand

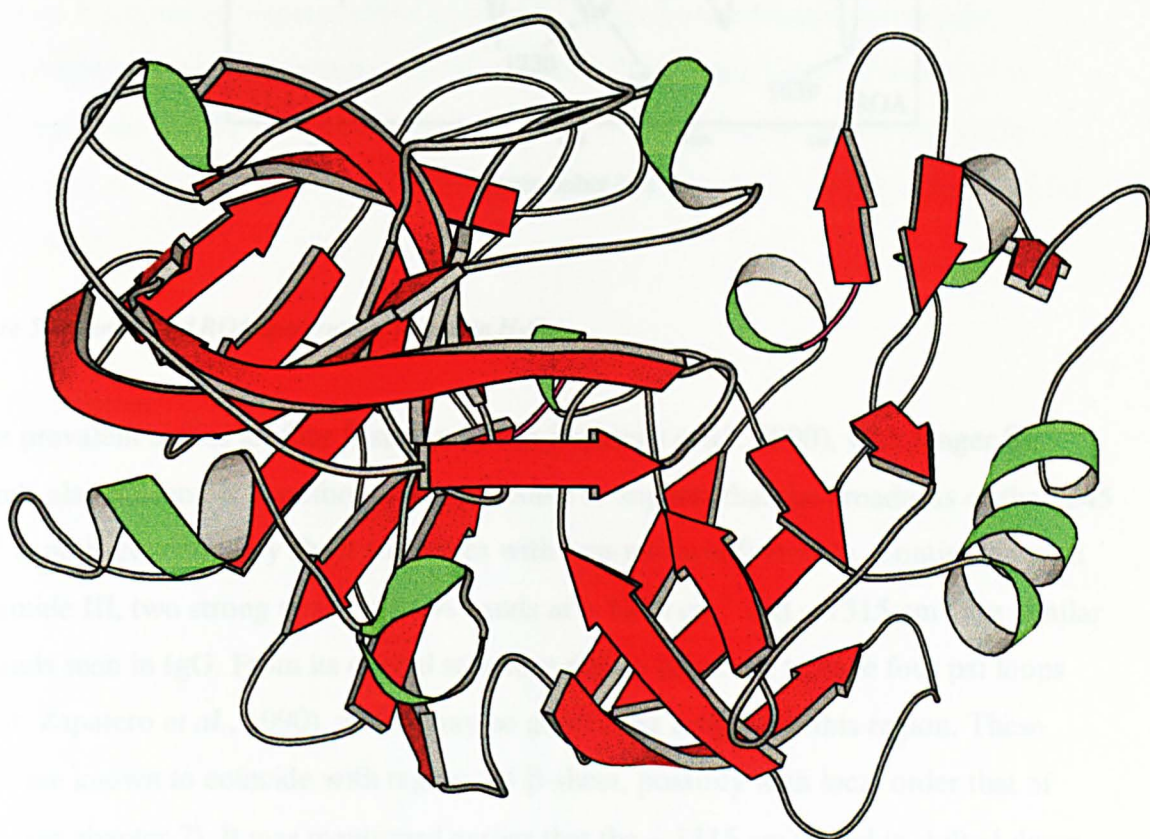


Figure 5.7 MOLSCRIPT diagram of pepsin, 3pep.

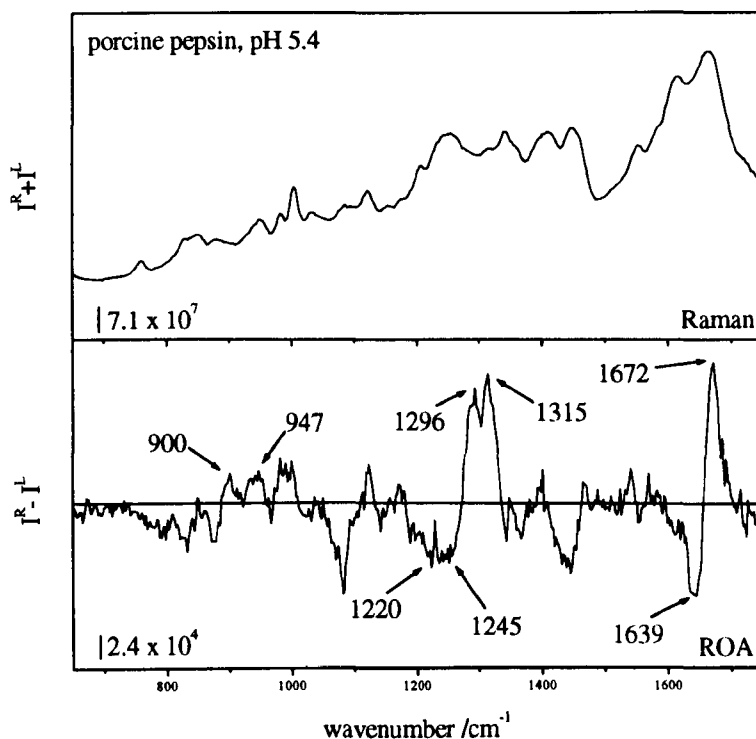


Figure 5.8 Raman and ROA spectrum of pepsin in H_2O .

quite prevalent across all four β -sheets (Adab-Zapatero *et al.*, 1990), with longer β -strands also present. It therefore seems plausible to suggest that the broadness of the 1245 cm^{-1} band is generated by sheet structures with non-uniform β -strands. Continuing with the amide III, two strong positive ROA bands at $\sim 1296\text{ }cm^{-1}$ and $\sim 1315\text{ }cm^{-1}$ are similar to bands seen in IgG. From its crystal structure pepsin is known to have four psi loops (Adab-Zapatero *et al.*, 1990), which may be generating a signal in this region. These loops are known to coincide with regions of β -sheet, possibly with local order that of PPII (see chapter 7). It was mentioned earlier that the $\sim 1315\text{ }cm^{-1}$ band is shifted down by $\sim 7\text{ }cm^{-1}$ when the H_2O spectrum of IgG was compared with avidin. Tentatively, therefore, it is suggested that this band is monitoring delicate perturbations in the local (ϕ, ψ) propensities of amino acid residues in extended, disordered polypeptide loops (see chapter 7). Also, the 1296 cm^{-1} band is in a region that may be viewed as a superposition of α -helix and β -turn structure. However, a previous D_2O ROA study by Wilson (1996) suggests that this band originates in mostly α -helix vibrations.

In the amide I of the ROA spectrum the usual negative/positive couplet is present. Here, the positive peak is located at $\sim 1672\text{ cm}^{-1}$, as expected for a β -sheet protein. However, the negative peak is centred at $\sim 1639\text{ cm}^{-1}$, which is similar to avidin, but differs from IgG by $\sim 10\text{ cm}^{-1}$, with the centre of the couplet at $\sim 1656\text{ cm}^{-1}$. In IgG and avidin both of their amide I couplets were centred at $\sim 1662\text{ cm}^{-1}$. This 6 cm^{-1} down shift is possibly reflecting the fact that pepsin has some ($\sim 10\%$) α -helix. Finally, the presence of two small side bands at $\sim 1685\text{ cm}^{-1}$ and $\sim 1692\text{ cm}^{-1}$ in the H_2O ROA spectrum is rather intriguing, as Krimm has suggested that this region is representative of β -turn structure (Krimm & Bandekar, 1986), demonstrating that the amide I of the ROA spectrum is a complex superposition of spectral bands from different polypeptide conformations. A rigorous mathematical analysis of the amide I and III region of the ROA spectrum should provide quantitative information on the secondary and tertiary structure of proteins. A more complete, higher quality data set is, however, required for such a study.

5.4. Discussion

Each of the β -sheet proteins in this chapter appears to have a large amount of disordered structure associated with regular β -strands. The 1309 cm^{-1} to 1315 cm^{-1} bands found in all three ROA spectra is now thought to originate in PPII loop structure ($\sim 1315\text{ cm}^{-1}$) and disordered structure with α -helix propensities ($\sim 1309\text{ cm}^{-1}$) (Barron *et al.*, in press) whereas, before they were thought to be signatures of β -sheet ($\sim 1315\text{ cm}^{-1}$) and canonical α -helix ($\sim 1309\text{ cm}^{-1}$). The opposite is true of the β -sheet 1238 cm^{-1} to 1245 cm^{-1} ROA band which was formerly assigned as PPII loop structure (Barron *et al.*, in press). Also, in pepsin the 1220 cm^{-1} ROA band is possible associated with hydrated β -sheet structure (see chapter 8) or β -turns. In addition, other studies were conducted (data not shown). For example, a deglycosylated variant of avidin was studied to see if deglycosylation had any effect on the polypeptide structure of this biomolecule. Unfortunately, this study could not be performed due to the poor solubility of this sample (Livnah *et al.*, 1993). Also, studies on molten globule substates of the $\text{C}_{\text{H}2}$ domain of IgG

(Vlasov *et al.*, 1996) were conducted (data not shown), but revealed little or no change in the ROA spectrum. Also, a study of polyclonal *versus* monoclonal IgG (provided by Glaxo) was conducted to see if differences occurred between these molecules in their ROA spectra. In addition, a similar ROA study of bovine IgG and human IgG was conducted. Both of these studies showed that little or no differences could be seen in comparing these samples.

Chapter 6

Serine Proteinase Inhibitors

Proteinase inhibitors are ubiquitous and varied with many different class-specific reactive sites (Laskowski & Kato, 1980). For example, carboxyl-, metallo- and sulphhydryl-proteinase all have inhibitors that are specific to each respective class of proteinase.

6.1. Serine Proteinase Inhibitors (serpins)

The serine proteinase inhibitors (serpins) is one such inhibitor family (Laskowski & Kato, 1980), and have evolved over the past ~500 to 1,000 million years, from the time of the first multicellular organisms (Huber & Carrell, 1989). Ovomucoids (chicken and turkey), an active plasma serpin (recombinant α_1 -antiproteinase) and an inactive serpin (ovalbumin) are examined in this chapter. Their physiological function is the prevention of unwanted proteolysis (Laskowski & Kato, 1980), which refers to the cleavage of proteins at specific polypeptide backbone sites. Several families of serine proteinase inhibitors exist, which are not all homologous, and have arisen by convergent rather than divergent evolution (Laskowski & Kato, 1980).

Each of the serpins listed in table 6.1 function by having an “active” loop, which docks with and mimics the transition state of its substrate (Perutz, 1992), to form a stable proteinase–proteinase-inhibitor complex. For example, trypsin and pancreatic trypsin inhibitor (pancreatic is a misnomer; this inhibitor can be found in other organs (Perutz, 1992)) form a complex with a dissociation constant of $\sim 10^{-13}$ and has a chemical half-life of \sim six weeks (Perutz, 1992). The formation of this complex is in some ways reminiscent of an antibody complexing with an antigen, with a large buried surface area ($\sim 1400 \text{ \AA}^2$) of the same order as an antibody-antigen complex (Perutz, 1992).

Table 6.1 Some characteristics of the serine proteinase inhibitors studied in this chapter.

Family	Mass (Da)	Glycosylated	Reactive site
Ovomucoids (chicken and turkey)	~ 25000	Yes (~ 25%)	Single (chicken), Double (turkey) & Triple (duck) Headed
Serpins (plasma) <u>serine</u> <u>proteinase</u> <u>inhibitors</u>	~ 42000 to 60000	Yes (variable, up to ~ 20%)	Single Headed, with some inactive forms such as ovalbumin

It can also be seen from table 6.1 that each of the given serpins are described as being single, double or triple headed (Laskowski & Kato, 1980), and refers to the fact that multiple reactive sites can exist on a single polypeptide chain. Multiheadedness is commonly achieved by gene duplication, with lima bean inhibitor from the Bowman-Birk family (a Bowman-Birk protein will be examined in the next chapter) the first to be discovered. The multiheadedness of serpins stands in contrast to other biologically active proteins where only single reactive sites are usually encountered (Laskowski & Kato, 1980). As well as providing multiple reactive sites, gene duplication also increases the molecular mass of the inhibitor, meaning that excretion of these molecules is prohibited (Laskowski & Kato, 1980). Interestingly, the plasma serpins (e.g. α_1 -antiproteinase) are all single headed (Laskowski & Kato, 1980). There is no clear reason for this. Multiheadedness also means that different enzymes can compete for the same inhibitor (Laskowski & Kato, 1980). For example, the ovomucoids are excellent inhibitors of both α -chymotrypsin and trypsin, with near equal activity for both samples (Wuyuan *et al.*, 1997). A different situation exists for α_1 -antiproteinase, with elastase and trypsin both competing for a single inhibitor site (Huber & Carrell, 1989).

6.1.1. The Active Loop

The canonical structure of the active loop in serpins is controversial. For example, α_1 -antitrypsin has a number of loop structure variants (Perutz, 1992; Elliott *et al.*, 1998). In early work the active loop was thought to be α -helical, based on a model of ovalbumin (an inactive serpin), as described by Perutz (1992), whereas more recent work by Elliott describes this structure as a twisted β -strand (Elliott *et al.*, 1998). X-ray studies on the structure of serine proteinase binding cavities show that a range of inhibitor active loop structures can be accommodated (Perutz, 1992). The variability of the active loop in α_1 -antitrypsin suggests that amino acid substitutions (single or multiple substitutions) are useful in controlling the activity of this inhibitor (Elliott *et al.*, 1998).

In mammalian serine proteinase the active site (binding pocket) is located between two β -sheet domains containing a serine, histidine, and a buried aspartate residue linked by hydrogen bonds. The histidine residue is polarised by the ionised aspartate residue, resulting in the polarisation of the serine hydroxyl group, rendering it nucleophilic (Perutz, 1992). An account of the catalytic details involved in this mechanism is given by Perutz (1992). The specificity of serine proteinase for particular inhibitors is controlled by a pocket adjacent to the active site (Perutz, 1992). In elastase, a mainly hydrophobic pocket is specific for short aliphatic chains; trypsin has an aspartate residue at the base of this pocket and cleaves bonds following basic residues; and chymotrypsin has a deep hydrophobic pocket and cleaves bonds following aromatic residues (Perutz, 1992).

A subtle physiological balance between serine proteinase and their inhibitors is maintained to ensure that the right amount of proteolysis is achieved (Perutz, 1992). If this balance is disrupted, chronic diseases may ensue (Perutz, 1992). A brief outline of these problems is given in the next section.

6.1.2. Some Medical Conditions Associated with Dysfunctional Serpins

It was mentioned in the previous section that amino acid substitutions (mutations) are implicated in modifying the inhibitory behaviour of serpin protease inhibitors. A good example of this is α_1 -antitrypsin. In the active loop of α_1 -antitrypsin a hydrophobic bond between methionine (Met173) and serine (Ser209) is easily cleaved (Poncz, 1992; Lohmann et al., 1994). A large conformational change in the structure of α_1 -

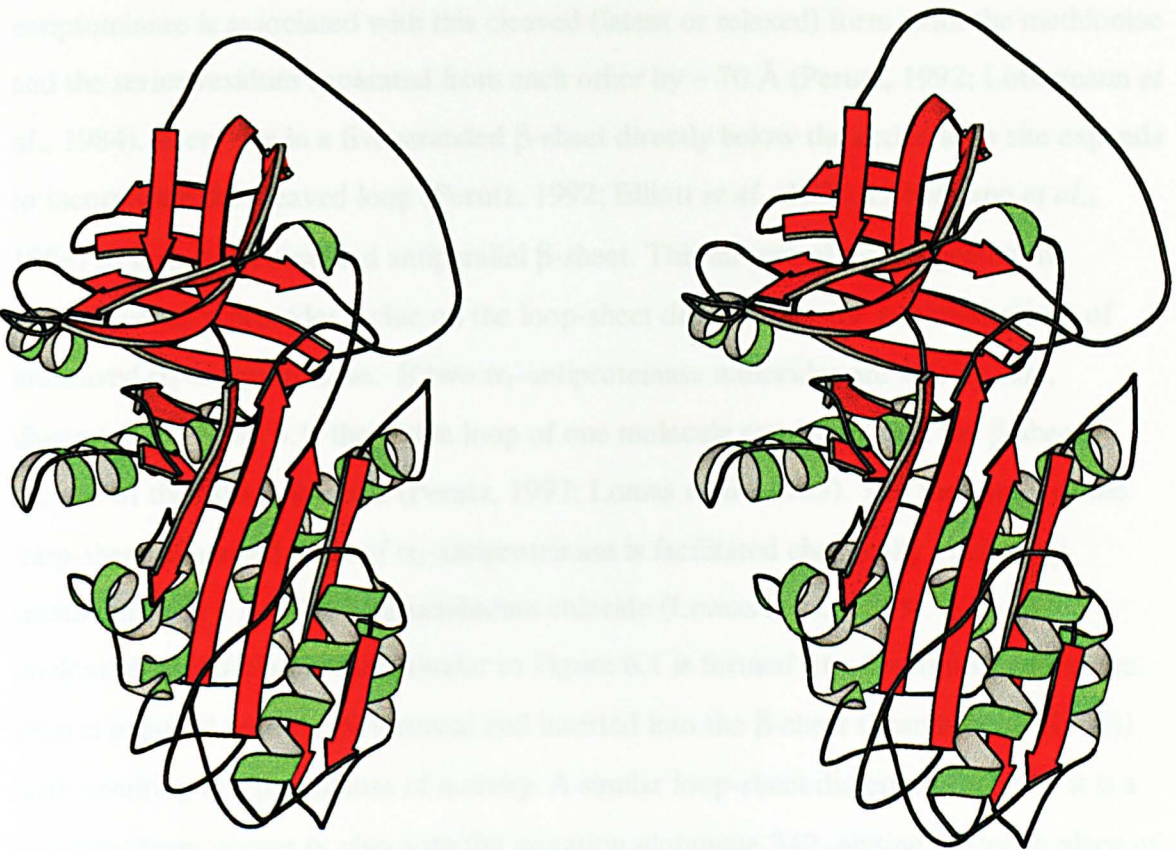


Figure 6.1 Loop-sheet dimerisation mechanism of α_1 -antitrypsin, Ipsi.

6.1.2. Some Medical Conditions Associated with Dysfunctional Serpins

It was mentioned in the previous section that amino acid substitutions (mutations) are implicated in modifying the inhibitory behaviour of serine proteinase inhibitors. A good example of this is α_1 -antiproteinase. In the active loop of α_1 -antiproteinase a scissile bond between methionine (Met358) and serine (Ser359) is easily cleaved (Perutz, 1992; Lobermann *et al.*, 1984). A large conformational change in the structure of α_1 -antiproteinase is associated with this cleaved (latent or relaxed) form, with the methionine and the serine residues separated from each other by $\sim 70 \text{ \AA}$ (Perutz, 1992; Lobermann *et al.*, 1984). A crevice in a five stranded β -sheet directly below the active loop site expands to incorporate the cleaved loop (Perutz, 1992; Elliott *et al.*, 1998; Lobermann *et al.*, 1984) to form a six stranded antiparallel β -sheet. This structural rearrangement is informative as it provides a clue on the loop-sheet dimerisation (or polymerisation) of uncleaved α_1 -antiproteinase. If two α_1 -antiproteinase molecules are side by side, illustrated in Figure 6.1, the active loop of one molecule can 'snap' into the β -sheet crevice of the other molecule (Perutz, 1992; Lomas *et al.*, 1995). The formation of the loop-sheet dimerised form of α_1 -antiproteinase is facilitated chemically *in vitro* by treatment with 1.1 or 1.5 M guanidinium chloride (Lomas *et al.*, 1995). Also, an uncleaved latent/relaxed form similar to Figure 6.1 is formed (the C-terminal end of the loop is exposed with the N-terminal end inserted into the β -sheet (Lomas *et al.*, 1995)), both resulting in a partial loss of activity. A similar loop-sheet dimerised (actually it is a polymer) form occurs *in vivo* with the mutation glutamate 342 \rightarrow lysine (lysine in place of glutamate) (Perutz, 1992; Lobermann *et al.*, 1984), and is a result of mutant α_1 -antiproteinase genes in homozygous carriers (Perutz, 1992). For example, the above substitution results in emphysema of the lung, with a concurrent reduction in α_1 -antiproteinase secretions (\sim one sixth of the normal level). In addition, sufferers of this mutation are predisposed to an associated cirrhosis of the liver (Perutz, 1992), due to inclusion body formation in hepatocytes from α_1 -antiproteinase aggregates. Physiologically this polymerisation reaction may be triggered by fever (Perutz, 1992). A relatively small change in body temperature from 37° to 41° C catalyses the

polymerisation of α_1 -antiproteinase (Perutz, 1992) (also used *in vitro* to produce polymers (Lomas *et al.*, 1995)). Consequently, new-born babies suffering from liver disease are monitored for signs of fever, as a means of controlling liver damage (Perutz, 1992).

As well as causing damage to liver and lung tissues, the malfunction of other serpins such as α_1 -antichymotrypsin are implicated in numerous pathologies of the nervous system (Adachi, 1990; Roa *et al.*, 1990; Tanzi, 1990; Sawaya, 1990; Abraham & Potter, 1990). These conditions are usually characterised by a build-up of amyloid deposits (plaques) consisting of insoluble fibrils ~ 6 to 10 nm in length (Frautschy *et al.*, 1998; Kusumoto *et al.*, 1998), which are characterised by having an extensive anti-parallel pleated β -sheet structure (Frautschy *et al.*, 1998; Kusumoto *et al.*, 1998). Some examples are (Adachi, 1990; Roa *et al.*, 1990; Tanzi, 1990; Sawaya, 1990; Abraham & Potter, 1990):

- The normal ageing process of the brain;
- Alzheimer's disease;
- Down's syndrome.

In Alzheimer's disease, for example, α_1 -antichymotrypsin is tightly associated, through a coinfection process (Frautschy *et al.*, 1998), with a 40 to 42 amino acid peptide known as amyloid β -protein (A β) (Frautschy *et al.*, 1998), which is the primary cause of plaque deposition (Frautschy *et al.*, 1998; Kusumoto *et al.*, 1998). The formation of this plaque follows first order kinetics and is thought to be initiated by α_1 -antichymotrypsin forming a complex with A β through a β -strand insertion of the N-terminus of the A β peptide (Frautschy *et al.*, 1998; Kusumoto *et al.*, 1998). The exact details of this activation mechanism are still a matter of debate (Frautschy *et al.*, 1998).

The deleterious effects of plaque formation can, however, be offset by the use of dimethyl sulfoxide (DMSO) in some conditions (Adachi, 1990) and has been used with some success for almost twenty years, when administered orally. It is thought to act on the precursors of amyloid deposits, rather than acting directly on existing amyloid substances (Adachi, 1990).

6.2. Chicken and Turkey Ovomucoid

A major constituent of avian egg whites are the ovomucoids (Matsuda *et al.*, 1981), illustrated in Figure 6.2 (only one domain is shown). Each of the ovomucoids have a tandem, three domain structure and are glycoproteins, with a carbohydrate content of ~ 25 % (Watanabe *et al.*, 1981), a molecular weight of ~ 25 kDa and 185 amino acid residues (Matsuda *et al.*, 1981). It was mentioned earlier in this chapter that the ovomucoids are multiheaded (table 6.1), which was regarded with surprise when first discovered (Laskowski & Kato, 1980). Also, the connecting peptide between the second and third domain is easily hydrolysed (Laskowski & Kato, 1980), resulting in an active domain III and the double domain I-II. However, although all of the serine proteinase inhibitors are characterised by having an exposed active loop, the ovomucoids differ from α_1 -antiproteinase by not having a latent form (Fujinaga *et al.*, 1987). This is also characteristic of the Bowman-Birk inhibitor (Werner & Wemmer, 1992) and ovalbumin (Bruch *et al.*, 1988).

In the previous chapter it was mentioned that, as a general rule, the region of the ROA spectrum from ~ 600 to 850 cm^{-1} is of little diagnostic value. However, in chicken ovomucoid two positive, weak ROA bands are seen at ~ 665 cm^{-1} and 810 cm^{-1} , respectively (see Figure 6.3). Other proteins show these bands but they are usually masked by a strong baseline distortion starting from ~ 800 cm^{-1} down to lower wavenumber. The 665 cm^{-1} band is almost certainly a C-S stretch (Tu, 1986), and interestingly appears only very weakly in the Raman spectrum. In the D_2O spectrum the 665 cm^{-1} band is retained and suggests that this mode is highly localised, with little or no coupling to vibrational modes involving exchangeable protons. The ROA band at ~ 810 cm^{-1} is in a region generally assigned to tryptophan or tyrosine residues (Tu, 1986). Similar bands also appear in the ROA spectrum of turkey ovomucoid (see Figure 6.4).

The backbone skeletal stretch regions of the chicken and turkey ovomucoid ROA spectra, however, differ from each other. In chicken ovomucoid two positive ROA bands appear at ~ 926 cm^{-1} and 967 cm^{-1} , and indicate the presence of α -helix (Tu, 1986; Wilson, 1996; Barron & Hecht, 1994). In turkey ovomucoid similar bands appear, but are weaker,

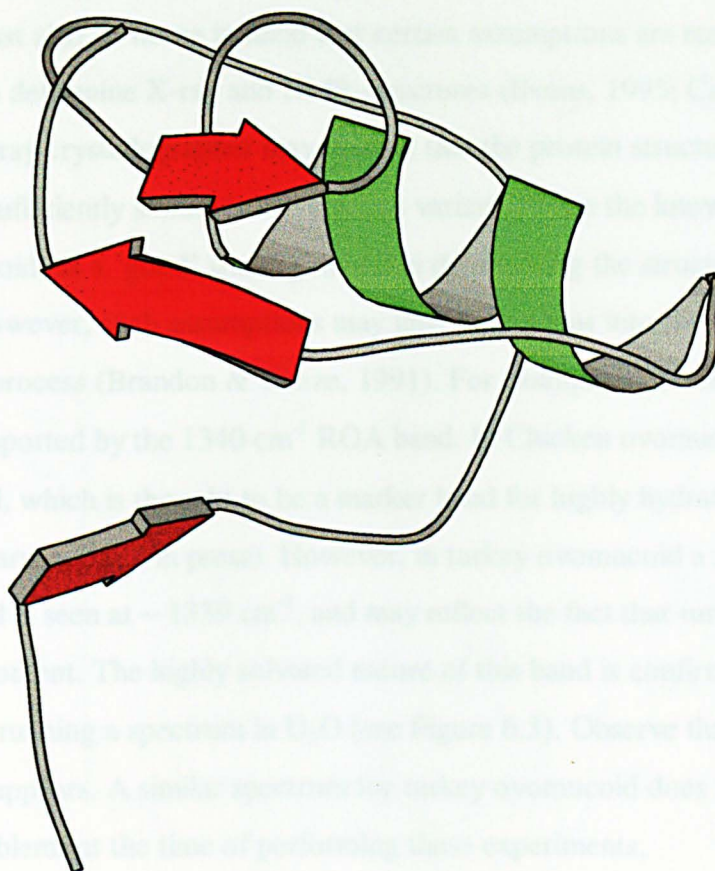


Figure 6.2 MOLSCRIPT diagram of Japanese quail ovomucoid (only third domain is shown), Iovo.

and may indicate that chicken ovomucoid has a greater α -helix content. X-ray and NMR studies on fragments of turkey and silver pheasant ovomucoid indicate an α -helix content of $\sim 17\%$ (Bode & Epp, 1985; Hoogstraten *et al.*, 1995). No X-ray or NMR studies exist on chicken ovomucoid that could confirm if ROA is detecting differences in the α -helix content. In comparing proteins from such closely related species an opportunity exists to test the reliability of ROA as a tool for structure elucidation. If X-ray and NMR studies showed that chicken and turkey ovomucoid were substantially similar in their secondary structure content, then a rethink of ROA spectroscopic interpretations may be needed. However, it must also be borne in mind that certain assumptions are made in the models that are used to determine X-ray and NMR structures (Evans, 1995; Carter, 1998). For example, an X-ray crystallographer may assume that the protein structure of turkey ovomucoid is sufficiently similar to the chicken variant, to use the known structure of turkey ovomucoid, as a 'good' starting model in determining the structure of chicken ovomucoid. However, such assumptions may introduce a bias into the structure determination process (Brandon & Tooze, 1991). For example, differences in the α -helix content are supported by the 1340 cm^{-1} ROA band. In Chicken ovomucoid a positive $\sim 1340\text{ cm}^{-1}$ band, which is thought to be a marker band for highly hydrated α -helix, is clearly seen (Barron *et al.*, in press). However, in turkey ovomucoid a similar, but less prominent band is seen at $\sim 1339\text{ cm}^{-1}$, and may reflect the fact that turkey ovomucoid has a lower helix content. The highly solvated nature of this band is confirmed in chicken ovomucoid by running a spectrum in D_2O (see Figure 6.3). Observe that this band completely disappears. A similar spectrum for turkey ovomucoid does not exist due to instrument problems at the time of performing these experiments.

As well as having an α -helix content of $\sim 17\%$, the ovomucoids are known to have a variable amount of β -sheet (Bode & Epp, 1985; Carter, 1998; Hoogstraten *et al.*, 1995; Weber *et al.*, 1982), with a lower β -sheet limit of $\sim 10\%$ (turkey (Hoogstraten *et al.*, 1995)) and an upper limit of $\sim 21\%$ (Japanese quail (Weber *et al.*, 1982)). In the ROA spectrum of chicken ovalbumin (Figure 6.3) a positive band $\sim 1024\text{ cm}^{-1}$ suggests the presence of right-twisted β -strands (Wilson, 1996). A similar band appears in the ROA

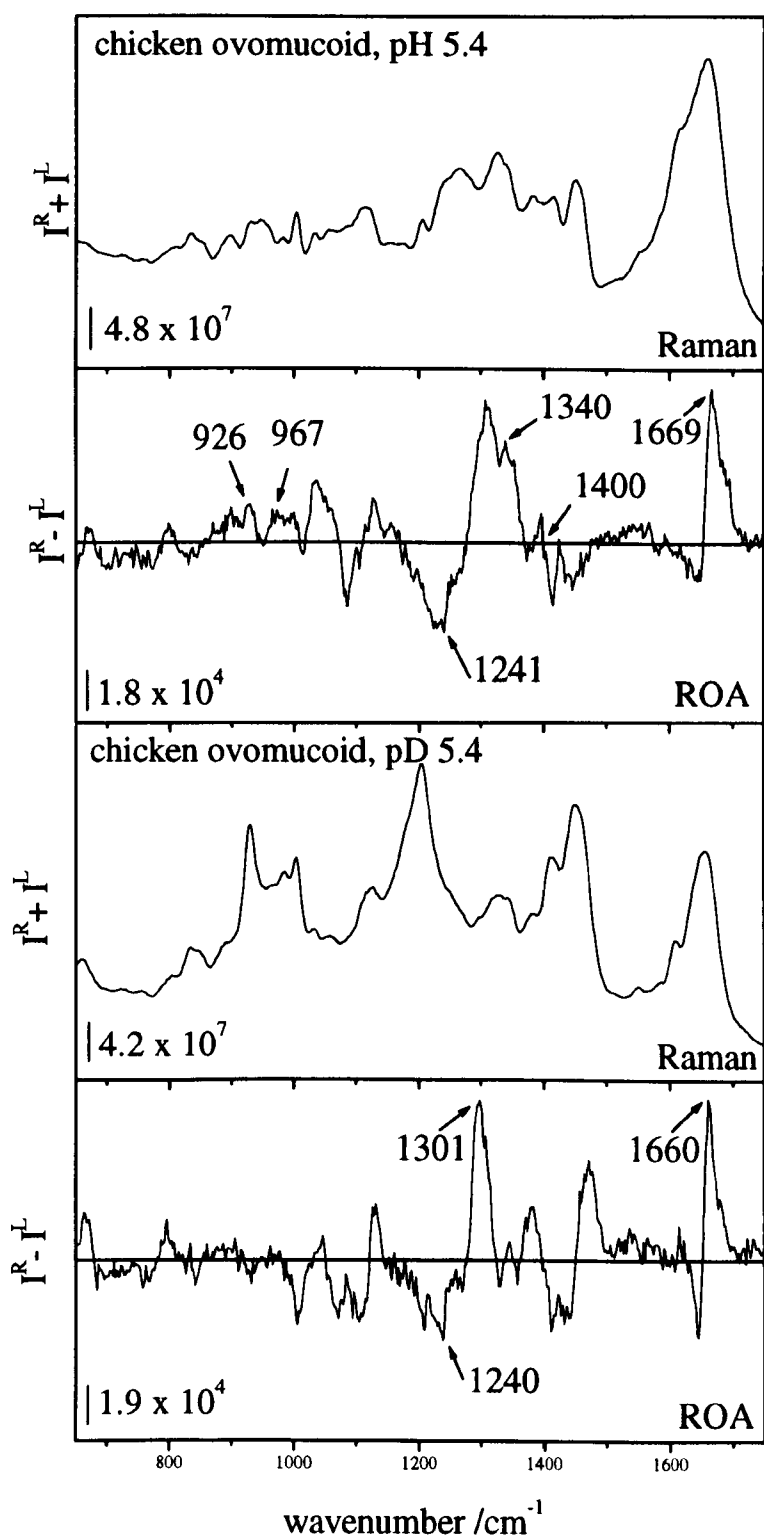


Figure 6.3 Raman and ROA spectra of chicken ovomucoid in H_2O (top) and D_2O (bottom).

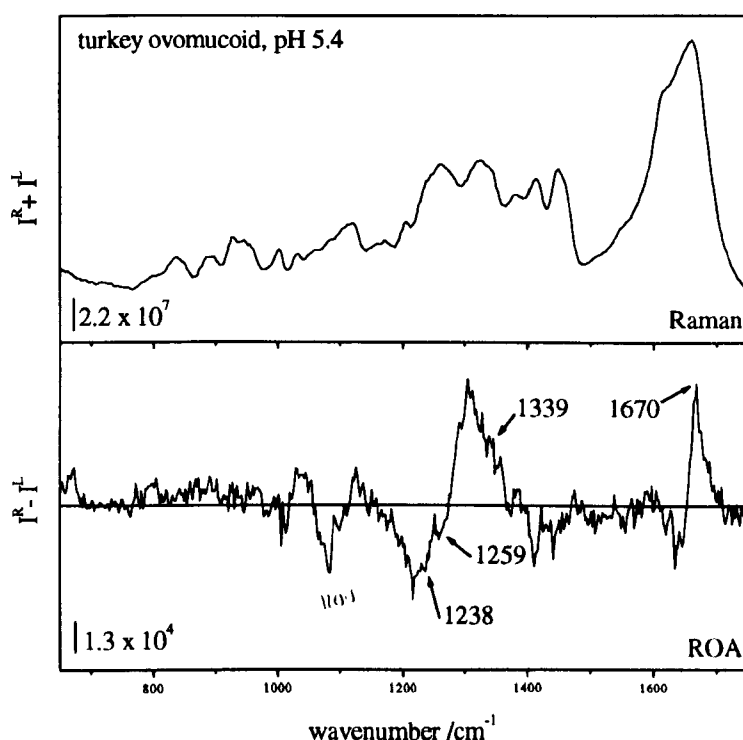


Figure 6.4 Raman and ROA spectra of turkey ovomucoid.

spectrum of turkey ovomucoid, but is slightly weaker, which may suggest a difference in the β -sheet content of these proteins. In the D_2O spectrum of chicken ovomucoid (Figure 6.3) the 1024 cm^{-1} band disappears and suggests that the β -strands are solvated. The MOLSCRIPT diagram of Japanese quail ovomucoid (see Figure 6.2) does suggest a rather open structure with the three β -strands located on the surface of the protein fragment (domain III). Other β -sheet features include the negative band at $\sim 1241\text{ cm}^{-1}$ for chicken ovomucoid and 1238 cm^{-1} for turkey ovomucoid. Both of these bands have a similar appearance making it difficult to discriminate between both proteins in this region of the ROA spectrum. However, in the ROA spectrum of turkey ovomucoid a negative band at $\sim 1259\text{ cm}^{-1}$ is clearly resolved, which is sometimes assigned as β -turn from Raman studies (Krimm & Bandekar, 1980, 1986), and possibly indicates a difference in the structures of turkey and chicken ovomucoid. In chicken ovomucoid this band is a weak shoulder and not well resolved. Deuteration of the 1240 cm^{-1} band results in a greater loss of intensity at high wavenumbers (above 1250 cm^{-1}) in the turn region of the ROA spectrum. In the

amide I region both ROA spectra have positive, strong bands at $\sim 1669\text{ cm}^{-1}$ (chicken) and 1670 cm^{-1} (turkey), indicating the presence of β -sheet. Qualitatively, both bands appear to be different, with the positive 1669 cm^{-1} band of chicken ovomucoid appearing to be a little broader than the same band in turkey ovomucoid. In chicken ovomucoid three very weak shoulders at $\sim 1685\text{ cm}^{-1}$, 1690 cm^{-1} and 1695 cm^{-1} are in a region assigned as β -turn (Krimm, 1987; Krimm & Bandekar, 1986). A similar weak shoulder is seen in turkey at $\sim 1690\text{ cm}^{-1}$, but is less prominent. Deuteration effects in the amide I of the ROA spectrum are quite strong, with the positive 1669 cm^{-1} band in ovomucoid shifted to 1660 cm^{-1} . The position of the amide I band is in a region that borders the α -helix and β -sheet regions (see previous chapter), which fits with the X-ray crystal structures and NMR (Bode & Epp, 1985; Carter, 1998; Hoogstraten *et al.*, 1995; Weber *et al.*, 1982).

In both chicken and turkey ovomucoid the carbohydrate content is estimated at $\sim 25\%$. However, little or no bands appear to originate from the carbohydrate. One possible exception is the sharp positive-negative band in chicken centred at $\sim 1400\text{ cm}^{-1}$, which has been assigned as a $\beta(1-3)$ glycosidic band in orosomucoid (Bell, 1994). In chapter four it was mentioned that this apparent lack of carbohydrate bands in glycoproteins is quite common.

6.3. Ovalbumin and Recombinant α_1 -Antiproteinase

Ovalbumin, although a serpin, is inactive (Perutz, 1992) and is a major constituent of avian egg-white (Stein *et al.*, 1991). Its function is unknown, but does have a heat-induced active form known as S-ovalbumin (Herald & Smith, 1992; Mellet *et al.*, 1996). Ovalbumin is a tetrameric α/β glycoprotein (Stein *et al.*, 1991) with a low carbohydrate content of $\sim 2.5\%$ (Stein *et al.*, 1991) and a monomer mass of $\sim 45\text{ kDa}$, with 385 amino acid residues. The structure of ovalbumin is illustrated in Figure 6.5, with only chain A shown for convenience. In early X-ray crystallographic studies ovalbumin was used, although inactive, to model the loop site of its active analogues such as α_1 -antiproteinase (Perutz, 1992). In ovalbumin an α -helix is present in an equivalent position to the active

loop of α_1 -antiproteinase (Perutz, 1992). This led to speculation on the canonical structure of the active loop (Perutz, 1992). It was suggested, for example, that α_1 -antiproteinase would have a similar canonical active loop structure to ovalbumin (Perutz, 1992). Recent X-ray crystallographic studies, however, are at odds with this interpretation (Elliott *et al.*, 1998). Studies on non-glycosylated recombinant α_1 -antiproteinase have demonstrated that a right-twisted β -strand resides in the active site (Elliott *et al.*, 1998). The overall structure of α_1 -antiproteinase is however similar to ovalbumin, as can be seen from the MOLSCRIPT diagrams in Figures 6.1 and 6.5. Both proteins have comparable amounts of secondary structure, with 31.9% and 28.2% of β -strand for ovalbumin and α_1 -antiproteinase, respectively; and 29.5% and 26.1% of α -helix, respectively (Elliott *et al.*, 1998; Stein *et al.*, 1991).

In the H_2O ROA spectra of both ovalbumin and α_1 -antiproteinase shown in Figure 6.6 the usual α -helix signals are present at low wavenumber at $\sim 850\text{ cm}^{-1}$ to 950 cm^{-1} . Note that it was not possible to obtain ROA spectra of ovalbumin and α_1 -antiproteinase in D_2O , due to instrument problems. Next, ovalbumin and α_1 -antiproteinase both have a medium-strong, positive ROA band at $\sim 1124\text{ cm}^{-1}$ and 1125 cm^{-1} , respectively, and appear frequently in proteins with both α -helix and β -sheet structure (Wilson, 1996). The positive bands at $\sim 1344\text{ cm}^{-1}$ (ovalbumin) and $\sim 1342\text{ cm}^{-1}$ (α_1 -antiproteinase) are also indicative of α -helix. In a previous ROA study of the D_2O spectrum of ovalbumin by Wilson (1996) this band completely disappears, suggesting that it is acting as a marker band of α -helices that are more solvated than other α -helices from lower wavenumber bands. An alternative interpretation to this is that different vibrational modes within an individual α -helix have different degrees of sensitivity to the effects of solvation. That is, for example, that the polypeptide backbone vibrational modes ($\sim 850\text{ cm}^{-1}$ to $\sim 950\text{ cm}^{-1}$) arise from vibrations that mix only very slightly with vibrations from exchangeable amide protons. However, these modes do have a C-N stretch ($\sim 20\%$) and C-N-C $_{\alpha}$ deformations (Krimm & Bandekar, 1986); whereas, in the amide III region of the ROA spectrum a greater extent of mixing occurs with exchangeable protons. Normal-mode Raman calculations on α -poly(L-alanine), suggest that the 1340 cm^{-1} band has an in-plane

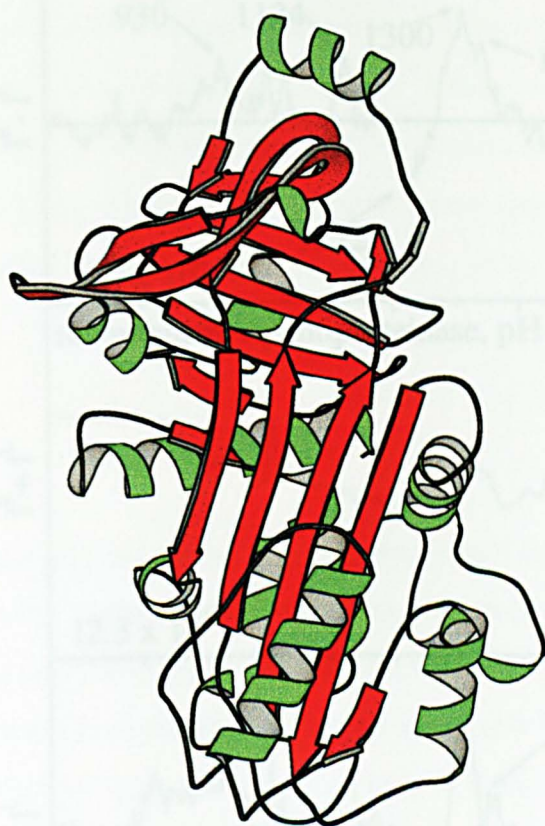


Figure 6.5 MOLSCRIPT diagram of ovalbumin (chain A), Iova.

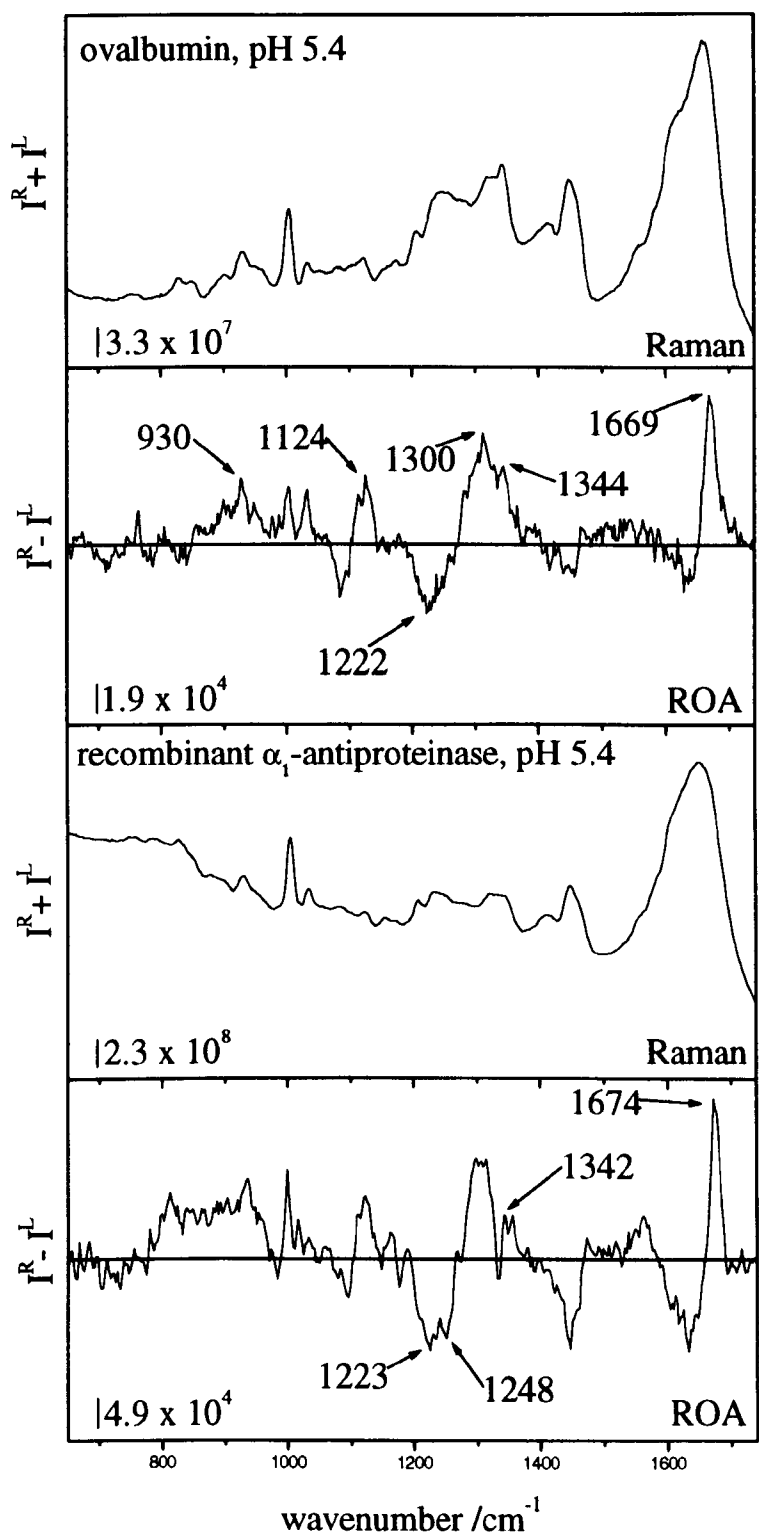


Figure 6.6 Raman and ROA spectra of ovalbumin in H_2O (top) and recombinant α_1 -antiproteinase in H_2O (bottom).

angle bend N-H mode which contributes ($\sim 12\%$) to the superposition of vibrations that make up this band. The 1300 cm^{-1} α -helix band has no such contribution (Krimm, 1987; Krimm & Bandekar, 1986).

The presence of β -strand structure in ovalbumin and α_1 -antiproteinase is alluded to, in the H_2O ROA spectra, by a broad, negative band at $\sim 1222\text{ cm}^{-1}$ in ovalbumin, with a weak, ill-resolved negative band at $\sim 1245\text{ cm}^{-1}$. The equivalent bands in α_1 -antiproteinase at $\sim 1223\text{ cm}^{-1}$ and $\sim 1248\text{ cm}^{-1}$ are better resolved. This may be an indicator of a more ordered β -strand structure in α_1 -antiproteinase. In the amide I the presence of β -strand is supported by the 1669 cm^{-1} band (ovalbumin) and the 1674 cm^{-1} band (α_1 -antiproteinase). In the H_2O spectra of ovalbumin and α_1 -antiproteinase the amide I bands differ in their line widths, with α_1 -antiproteinase appearing to be narrower, and possibly supports the claim that the β -strand structure of this protein is more ordered.

6.4. Discussion

It was hoped that the results of this chapter would be more informative, as there is much to be said of the serpins. Studies on the trypsin and chymotrypsin complexes of chicken and turkey ovomucoid (data not shown), which were conducted to determine any structural rearrangements of the inhibitors, were impossible to interpret, as a high fluorescence background was associated with each of these complexes. In addition, a study on a glycosylated variant of recombinant α_1 -antiproteinase (see chapter 8 for a brief discussion on this) appeared to give a prominent positive ROA band at $\sim 1340\text{ cm}^{-1}$ and suggested that glycosylation had a pronounced effect on the polypeptide structure of this biomolecule. Also, in the ROA spectra of chicken and turkey ovomucoid small differences appear at $\sim 1340\text{ cm}^{-1}$, $\sim 1259\text{ cm}^{-1}$ and in the backbone skeletal stretch region. For example, the 1340 cm^{-1} ROA band is thought to originate in hydrated α -helix (Barron *et al.*, in press) and shows that, although the turkey and chicken ovomucoid samples are from closely related species, differences in their polypeptide structures do occur.

Chapter 7

Proteins with Irregular Native Folds

The biological function of proteins is usually associated with a well-defined tertiary fold (Darby & Creighton, 1993). However, some exceptions to this maxim exist (Gast *et al.*, 1995). Phosvitin, invertase, Bowman-Birk proteinase inhibitor and metallothionein are a few of the exceptions, which are investigated in this chapter. In addition, homopolypeptides such as poly(L-lysine) and poly(L-glutamic acid), although not biologically active, are used to provide models of unordered 'random coils' (Wilson *et al.*, 1996b).

ROA reveals subtle differences in the disordered structures found in native proteins. It is proposed that two forms of disordered structure are revealed by ROA (Smyth *et al.*, to be published). Namely, 'unstructured' disorder, as revealed by phosvitin, and 'structured' disorder in the Bowman-Birk proteinase inhibitor together with metallothionein and invertase. Unstructured disorder appears to be dynamic and is similar to an ensemble of interconverting conformations found in a model random coil and reduced (unfolded) proteins such as lysozyme and ribonuclease A, where there is a distribution of Ramachandran (ϕ, ψ) angles corresponding to each amino acid residue. Structured disorder is static and is reminiscent of the structure found in loops, which contains a sequence of residues with fixed, non-repetitive Ramachandran (ϕ, ψ) angles.

One of the most interesting proteins with a native random coil conformation is prothymosin α (Gast *et al.*, 1995). No secondary structure is discernible under physiological conditions, but it does exhibit persistence of direction and curvature, as detected by NMR (Gast *et al.*, 1995). As is indicated by Gast *et al.* (1995), this finding raises the question of whether this is a common occurrence in biologically active biomolecules. The conditions under which prothymosin α functions is speculative. It has been suggested by Gast *et al.* (1995) that interactions with membranes, receptor molecules or chromatin induce a specific three-dimensional fold for the correct function of this biomolecule. An alternative suggestion, by the same authors, is that a nonfolded conformation may be advantageous, in the sense that a

nonfolded structure is free to interact with a greater range of target biomolecules. Prothymosin α has been implicated in the proliferation of cells found in cancer patients (Gast *et al.*, 1995). In addition, prothymosin α is implicated in the immunological response in the improvement of depressed killer-cells in immunodeficient patients. A second example of a random coil protein, found under physiological conditions, is the so-called τ -protein, which is found in sufferers of Alzheimer's disease (Gast *et al.*, 1995), and is similar in size to prothymosin α (both of these biomolecules have ~ 100 amino acid residues). Unfortunately, samples of these proteins were not available for ROA studies.

7.1. Unordered Poly(L-lysine) and Poly(L-glutamic acid)

The homopolypeptides of poly(L-lysine) (hydrobromide, $\overline{M}_w = 26000$ by viscosity) and poly(L-glutamic acid) (hydrobromide, $\overline{M}_w = 20000$ by viscosity) have a long history of being used as models for protein structure elucidation by CD spectroscopists (Davidson & Fasman, 1967; Barskaya & Ptitsin, 1971; Tiffany & Krimm, 1969; Holzwart & Doty, 1965; Greenfield & Fasman, 1969) and vibrational spectroscopists (Carey, 1982) since ~ 1965 . Theoretical studies by Krimm have also proved useful (Krimm & Bandekar, 1986), with the proviso that it is always kept in mind that these samples contain only one type of amino acid residue.

The Raman and ROA spectra of disordered poly(L-lysine) and poly(L-glutamic acid) in H_2O are presented in Figure 7.1. Here disorder is generated in the homopolypeptide samples by changes in pH conditions which produce charged side chains that repel each other (Bergethon, 1998). Early ROA studies of poly(L-lysine) ably demonstrated that when unordered homopolypeptide samples are compared to the ROA spectra of unfolded proteins such as hen lysozyme or ribonuclease A (Wilson, 1996; Wilson *et al.*, 1996b), the general appearance of these spectra is quite different to their reduced protein equivalents. In the backbone region of the ROA spectrum of disordered poly(L-lysine) $\sim 900\text{ cm}^{-1}$ to 970 cm^{-1} a weak negative band appears; whereas α -helical poly(L-lysine) has a positive ROA band in this region (Wilson, 1996; Wilson *et al.*, 1996b). Tiffany and Krimm (1968) proposed that unordered homopolypeptides possess locally-ordered regions of left-handed α -helix,

similar to PPII type structure (Wilson, 1996; Wilson *et al.*, 1996b), which gives a possible explanation for the origin of the 900 cm^{-1} to 970 cm^{-1} ROA band (Wilson, 1996; Wilson *et al.*, 1996b). VCD studies also appear to lend support to this interpretation (Keiderling, 1994). A similar feature also appears in the ROA spectrum of poly(L-glutamic acid). The extended amide III region of the poly(L-lysine) ROA spectrum is also informative. The absence of the positive $\sim 1340 \text{ cm}^{-1}$ band confirms that no hydrated α -helix is present in the samples of disordered poly(L-lysine) or poly(L-glutamic acid). Two positive peaks, for poly(L-lysine) only, at $\sim 1320 \text{ cm}^{-1}$ and 1300 cm^{-1} suggest that there is some propensity for locally ordered regions of β -type structure (possibly loop structure with PPII or β -turn propensities) and α -helix, respectively (Smyth *et al.*, to be published). The presence of an α -helix band at $\sim 1300 \text{ cm}^{-1}$, may seem at odds with the suggestion of the left-handed α -helix assignment at lower wavenumber, as the ROA signal from left-handed α -helix would be expected to cancel out the effects of the ROA signal from conventional right-handed α -helix (Wilson, 1996; Wilson *et al.*, 1996b). However, interchain hydrogen bonds are absent from the PPII helix (Wilson, 1996; Wilson *et al.*, 1996b; Creighton, 1994; Paterlini & Freeman, 1986); hydrogen bonds are also absent from this conformation in proteins. Here the PPII conformation is stabilised by hydrogen bonds to either water or side chains (Woody, 1992; Adzhubei *et al.*, 1987). Whereas, interchain hydrogen bonds of N-H groups have a significant influence on the contribution of N-H deformation coordinates to normal modes in the extended amide III region. Consequently, it is not expected that left-handed helix shows simple mirror-image versions of the α -helix ROA signatures here (Wilson, 1996; Wilson *et al.*, 1996b). The lack of β -turns in either sample is indicated by the absence of a negative band in the 1350 cm^{-1} to 1380 cm^{-1} region of the ROA spectrum (Wilson *et al.*, 1996b). A negative ROA band in poly(L-glutamic acid) is centred at $\sim 1400 \text{ cm}^{-1}$ and is possibly too high to be associated with turn structure. In the ROA spectrum of poly(L-lysine), at lower wavenumber, the negative band at $\sim 1240 \text{ cm}^{-1}$ in proteins (which is indicative of canonical β -strand) is shifted to slightly higher wavenumber ($\sim 1265 \text{ cm}^{-1}$), and is very weak, which suggests possibly that some lysine residues are clustering locally with Ramachandran (ϕ, ψ) angles in β -strand or turn regions. A second negative band located $\sim 1220 \text{ cm}^{-1}$ is also suggestive of residues with β -strand (ϕ, ψ) propensities (Smyth *et al.*, to be published). Poly(L-glutamic acid) differs

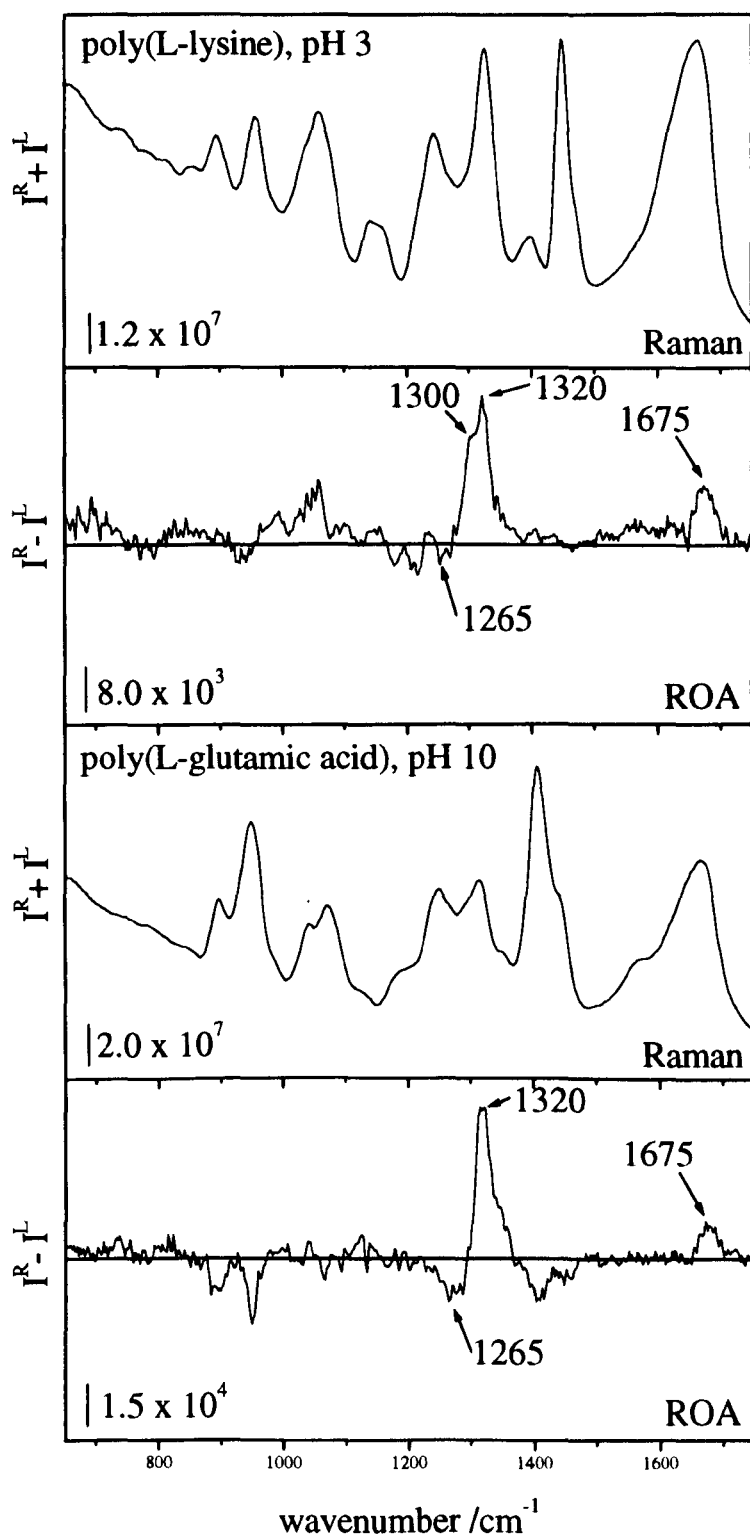


Figure 7.1 Raman and ROA spectra of poly(L-lysine) (top) and poly(L-glutamic acid) (bottom).

slightly, with only the negative 1265 cm^{-1} band appearing in this region of the ROA spectrum. This suggests that the appearance of the negative 1220 cm^{-1} ROA band may be dependent on the type of side chains present, at least in the case of homopolypeptides. In the amide I region of the ROA spectrum both samples have a positive peak at $\sim 1675\text{ cm}^{-1}$. In native protein samples a negative peak is usually associated (about $\sim 20\text{ cm}^{-1}$ lower down) with the higher wavenumber amide I band to form a couplet (see examples from other chapters), and is indicative of canonical secondary structure (Smyth *et al.*, to be published). Therefore, it is proposed that the absence of such a couplet is a marker of disordered structure.

7.2. Phosvitin

Hen egg phosvitin is a highly phosphorylated phosphoglycoprotein with a carbohydrate content of $\sim 6.5\%$ and a mass of $\sim 34\text{ kDa}$ (Burley *et al.*, 1989; Vogel, 1983; Prescott *et al.*, 1986; Byrne *et al.*, 1984), and is a major protein constituent of egg yolk (Burley *et al.*, 1989). Its amino acid composition sets it apart from other proteins, due to its high serine content (Burley *et al.*, 1989; Vogel, 1983; Prescott *et al.*, 1986; Byrne *et al.*, 1984). Phosvitin has ~ 216 amino acid residues, with ~ 123 serines, and the remainder being mostly lysine, histidine and arginine (Byrne *et al.*, 1989). Due to the unusual nature of phosvitin, it is hardly surprising that no X-ray crystallographic data exists on this protein. Spectroscopic evidence, however, does exist, but is contradictory in its determination of phosvitin's secondary structure content. Studies at neutral pH from NMR (Vogel, 1983), Raman (Prescott *et al.*, 1986) and VCD (Yasui *et al.*, 1990) all indicate the presence of an irregular structure, whereas FTIR (Lasso *et al.*, 1993) and UVCD (Yasui *et al.*, 1990) suggest some β -sheet as well as irregular structure. At a low pH of ~ 1.5 , β -sheet structure predominates (Prescott *et al.*, 1986). The low pH variant of phosvitin is, however, difficult to obtain in solution, due to gel formation (Prescott *et al.*, 1986).

The Raman and ROA spectrum of phosvitin are presented in Figure 7.2. It is immediately apparent that a large noise background plagues the ROA spectrum. However, a recent follow up studies on phosvitin with improved instrumentation, demonstrate that the results of this study are generally reliable. When the ROA

spectrum of phosvitin is compared with examples of folded proteins from previous chapters, it is obvious that significant differences exist between irregular random coil proteins and proteins with compact, well-defined structures. In the backbone region of the phosvitin ROA spectrum little or no bands are seen. Some very weak bands are discernible under the large noise background and may indicate the presence of some secondary structure. The usual negative–positive couplet found in the amide I region of the protein ROA spectra is missing, being replaced by a small positive peak $\sim 1675\text{ cm}^{-1}$, which is reminiscent of the homopolypeptides and some reduced (unfolded) proteins (Smyth *et al.*, to be published; Wilson, 1996; Wilson *et al.*, 1996b; Wilson *et al.*, 1996a). The origin of this 1675 cm^{-1} band is unknown, but may possibly originate in loop structure found in the β -sheet or PPII region of Ramachandran space (Barron *et al.*, in press). In the extended amide III region of the ROA spectrum a single, sharp negative-positive couplet crosses the baseline at $\sim 1290\text{ cm}^{-1}$, which is a little higher than normally found. In globular proteins this couplet is broader and is shifted to a lower wavenumber by ~ 10 to 20 cm^{-1} . The unordered homopolypeptides shown in the previous section both cross their respective baselines at $\sim 1275\text{ cm}^{-1}$. A negative ROA band at the low wavenumber end of the phosvitin amide III couplet is observed at $\sim 1267\text{ cm}^{-1}$, with a typical value of ~ 1240 to 1245 cm^{-1} in globular proteins, corresponding to β -strand structure. The 1267 cm^{-1} region is more characteristic of turn structure and borders the α -helix region of the ROA spectrum (Krimm & Bandekar, 1980). The positive part of the amide III couplet is located at $\sim 1318\text{ cm}^{-1}$ and is strongly suggestive of random coil with PPII propensities. However, recent results suggest that this band is a little lower down at $\sim 1312\text{ cm}^{-1}$ and is probably not indicative of PPII loop structure (Smyth *et al.*, to be published). The sharpness of the amide III couplet is also of interest. In reduced (unfolded) proteins this couplet band forms a moderately broad envelope and is suggestive of different amino acid residues clustering over a range of conformations. However, phosvitin would appear to have a more limited set of (ϕ, ψ) angle available to its amino acid residues (Smyth *et al.*, to be published). This may be due to the fact that phosvitin has a high concentration of phosphorylated serine residues. Finally, no sugar bands are visible in the ROA spectrum, due to the low carbohydrate content of phosvitin (Prescott *et al.*, 1986).

7.3. Invertase

Yeast external invertase is a glycoprotein with a molecular mass of ~ 270 kDa, comprising of two subunits, and a carbohydrate content of $\sim 50\%$ (Athés *et al.*, 1998). From the twelve glycosylation sites found in invertase, nine sites are always populated with high-mannose glycans (Athés *et al.*, 1998), with each glycan individually consisting of as many as ~ 20 mannose residues (Trimble & Maley, 1983). No X-ray or NMR data exists on the polypeptide structure of invertase, but some NMR data on the glycans found in invertase does exist (Trimble & Atkinson, 1986). As with phosvitin, the spectroscopic literature for invertase is contradictory. In a UVCD study of invertase's solution structure, it was concluded that little or no α -helix structure was present but that a significant amount of β -sheet structure predominates (Schülke & Schmid, 1988). Interestingly, theoretical structure prediction methods suggest that invertase belongs to a family of proteins with a β -propeller fold (Pons *et al.*, 1998). However, conventional Raman studies indicate that the solution structure of invertase is more likely to be α -helical, and that its solid state structure is a superposition of both α -helix and β -sheet (Athés *et al.*, 1998).

The Raman and ROA spectra of invertase are presented in Figure 7.2. In the amide I region of the ROA spectrum a small positive peak, reminiscent of a similar feature in phosvitin, is located ~ 1678 cm^{-1} . Consequently, it is suggested that little or no secondary structure is present in the solution structure of invertase. However, in the extended amide III region of the ROA spectrum there are two negative peaks at ~ 1245 cm^{-1} and 1222 cm^{-1} , which suggest that some β -sheet structure is present. In a recent Raman study of invertase the 1245 cm^{-1} band was assigned as disordered structure (Athés *et al.*, 1998). It is therefore possible that the 1245 cm^{-1} band is a superposition of both β -strands and disordered structure, with the disordered structure clustering in the β -sheet region of Ramachandran space. A positive ROA band at ~ 1312 cm^{-1} is borderline between α -helical structure and PPII loop structure and is of little help. A second positive ROA band ~ 1360 cm^{-1} is possibly due to tryptophan, as suggested by general Raman studies on proteins (Tu, 1986), although carbohydrate bands do occur in this region of the ROA spectrum (Bell, 1994). In fact, it is not implausible to suggest that the amide III region of the ROA spectrum originates in carbohydrate vibrations (Bell, 1994).

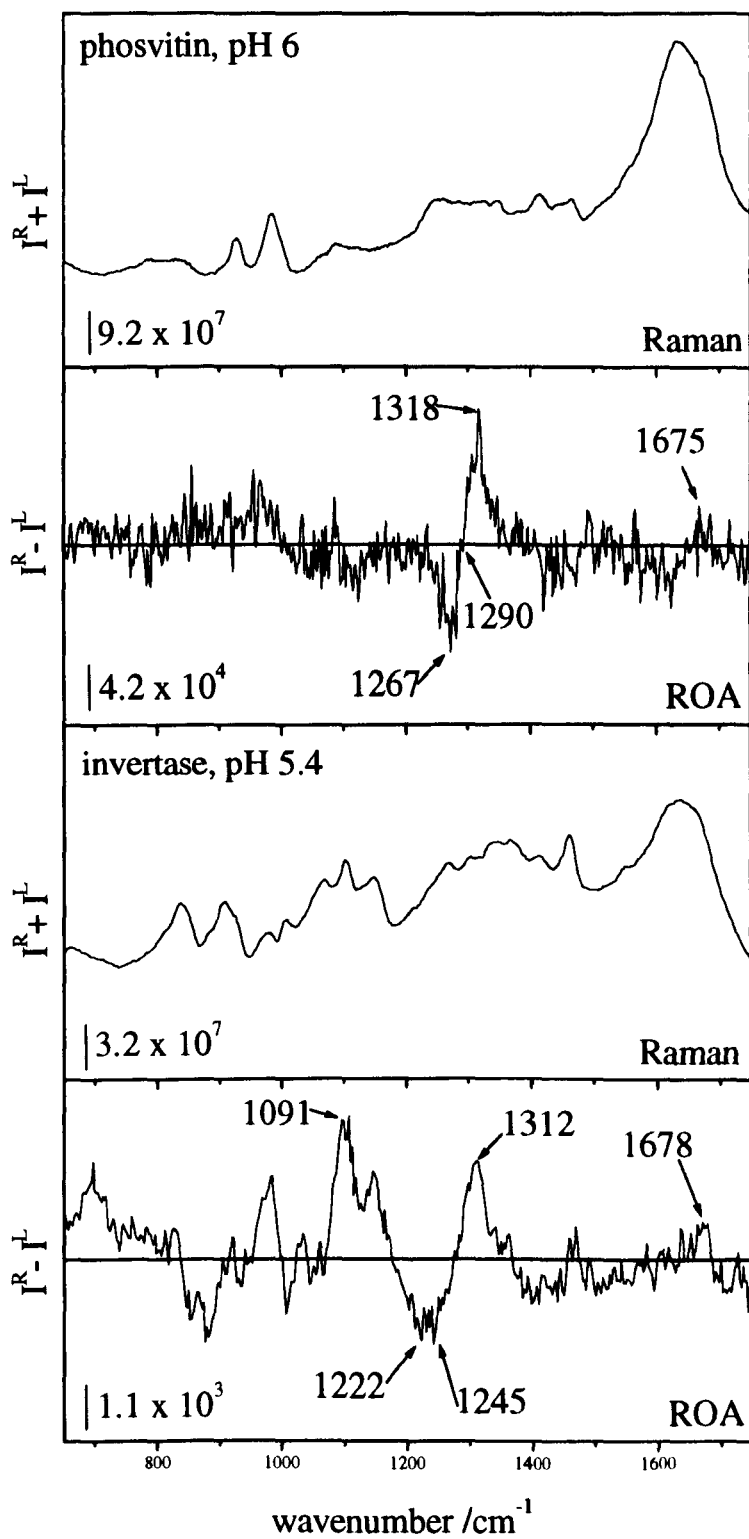


Figure 7.2 Raman and ROA spectra of phosvitin (top) and invertase (bottom).

Previous studies on small model carbohydrates reveal a rich diversity of vibrations in this region (Bell, 1994). Also, most of the bands below $\sim 1150\text{ cm}^{-1}$ look less like protein bands and more like carbohydrate bands. The ROA of carbohydrates demonstrates that the majority of ROA sugar bands are at low wavenumber, typically below $\sim 1200\text{ cm}^{-1}$ (Bell, 1994). However, it is not yet possible to make any reliable assignments of the carbohydrate ROA bands present in invertase. Conventional Raman spectroscopy has little success in making similar assignments (Athés *et al.*, 1998). There are two main reasons for this problem with regards to ROA. Firstly, there is the cost of buying model glycans. A typical 20 mg glycan sample is prohibitively expensive (\sim £1000 to £2000). Secondly, most of the glycans in invertase are large and have many linkage types (Trimble & Atkinson, 1986), both α and β with (1-2), (1-3) and (1-4) linkages, which may make the analysis of carbohydrate bands rather difficult to differentiate.

One interesting feature to note is the strong, positive 1091 cm^{-1} ROA band. A nearby band in the Raman spectrum $\sim 1081\text{ cm}^{-1}$ has been previously assigned to peripheral phosphates (PO_4), doubly ionised (Zwick *et al.*, 1989) oligosidic chains (Trimble & Maley, 1983). If the 1091 cm^{-1} ROA band is assigned similarly, it raises the interesting proposition that the phosphate groups in invertase are held in chiral conformations.

7.4. Bowman-Birk Proteinase Inhibitor

The Bowman-Birk family of proteinase inhibitors are a member of the serine proteinase inhibitors (Laskowski & Kato, 1980). A study of soybean Bowman-Birk proteinase inhibitor is presented in this chapter, rather than the previous chapter, as it is better considered as an example of a protein with an irregular fold as classified by the CATH structure database (Orengo *et al.*, 1997). Bowman-Birk inhibitors are small proteins with a molecular mass of $\sim 7\text{ kDa}$ and 71 amino acid residues (Werner & Wemmer, 1992), and are found in the seeds of leguminous plants (Laskowski & Kato, 1980). These proteins have an unusually high number of disulphide bridges that holds the polypeptide chain in a rigid, well-defined native fold consisting of two tandem homologous domains, each with a single proteinase binding site (Laskowski & Kato,

1980; de la Sierra *et al.*, 1999), with different specificities. One domain is specific to trypsin-like proteinase and the other is specific to chymotrypsin-like proteinase (de la Sierra *et al.*, 1999). Bowman-Birk inhibitors are currently of great interest, with application being found in the medical and agricultural sciences (de la Sierra *et al.*, 1999), resulting from their anti-carcinogenic properties (related to the chymotrypsin inhibitory activity), immune stimulating properties, and as insect repellents in plant defence systems (de la Sierra *et al.*, 1999; Ware *et al.*, 1997). Additionally, they are highly resistant to acid and heat attack. The MOLSCRIPT diagram of soybean Bowman-Birk proteinase inhibitor is illustrated in Figure 7.3. It comprises five short stretches of β -strand contained in two small sections of antiparallel sheet (Werner & Wemmer, 1992; de la Sierra *et al.*, 1999).

UVCD studies have been used to determine the secondary structure of the pea seed variant of the Bowman-Birk inhibitor by producing a best fit of the experimental CD spectrum against secondary structure parameters to give a structure content of 9% α -helix, 41% β -sheet, 22% β -turn and 28 % random coil (de la Sierra *et al.*, 1999). The solution structure of the soybean Bowman-Birk inhibitor has been studied by two-dimensional ^1H NMR and reveals that many amino acid residues are found in the PPII region of Ramachandran space (Werner & Wemmer, 1992). In addition, the chymotrypsin inhibitor domain is shown to be more disordered than the trypsin inhibitor domain, with a particularly high degree of disorder in the chymotrypsin inhibitor loop (Werner & Wemmer, 1992). However, the X-ray crystal structure of the Bowman-Birk inhibitor reveals a largely disordered structure, with some elements of β -sheet (de la Sierra *et al.*, 1999).

The Raman and ROA spectra of the Bowman-Birk inhibitor is presented in Figure 7.5. When the ROA spectrum of the Bowman-Birk inhibitor is compared with that of phosvitin it is qualitatively quite different. Sharper detailed structure is found throughout as in native proteins with regular folds. In the extended amide III of the ROA spectrum the negative 1240 cm^{-1} band is consistent with a significant amount of β -strand structure. The positive 1318 cm^{-1} band corresponds to the random coil region of the extended amide III and as remarked earlier a significant amount of PPII loop structure is found in the Bowman-Birk inhibitor, as revealed by NMR (Werner & Wemmer, 1992). Therefore, it would seem reasonable to suggest that PPII loops are synonymous with disordered structures. In the amide I region of the ROA spectrum a

strong, sharp positive peak at $\sim 1620\text{ cm}^{-1}$ is suggestive of a polypeptide structure with a significant amount of disordered structure, possibly originating in PPII loop structure. Turn structure is indicated by a negative peak at $\sim 1350\text{ cm}^{-1}$. Also, a negative peak at $\sim 930\text{ cm}^{-1}$ is similar to the band found in poly(L-lysine) and may originate from left-handed helix.

The overall appearance of the ROA spectrum gives the impression of a protein with elements of ordered and disordered structure with amino acid residues in well-defined regions of Ramachandran space, with, for example, a sharp negative ROA band at $\sim 1240\text{ cm}^{-1}$, which corresponds to the β -sheet found by UVCD (de la Sierra et al., 1999).

7.4. Metallothionein

Metallothioneins are a family of small proteins with a molecular mass of $\sim 6\text{ kDa}$ and have a high affinity for heavy metal ions such as Zn^{2+} and Cd^{2+} (Brown et al., 1992). They are believed to be involved in metal detoxification and metal homeostasis as well as specific metal-dependent functions against toxic metal ions (Brown et al., 1992). Each metallothionein contains ~ 61 amino acids and is composed of two cysteine-rich regions (Brown et al., 1992). There are ~ 20 genes encoding in the human genome, which are used to maintain the cellular level through seven distinct metal-binding sites (Brown et al., 1992) under the control of metal-responsive element (MRE) (Brown et al., 1992). The X-ray crystal structure of human metallothionein (Brown et al., 1991) and solution NMR structure (Kobayashi et al., 1996) of rat metallothionein are known and are similar (Brown et al., 1992). Also, the solution NMR structure of rabbit (Arancibia et al., 1988) and human metallothionein (Kobayashi et al., 1996) are known. Each variant contains two type I and II β -turns and one type III β -turn (Arancibia et al., 1988; Minamide et al., 1979), with varying degrees of disorder, which has $\sim 10\%$ α -helix (Minamide et al., 1979). The structure of metallothionein has been determined by X-ray crystallography (Kobayashi et al., 1996) and by NMR (Arancibia et al., 1988; Minamide et al., 1979). The MOLSCRIPT diagram of Bowman-Birk inhibitor, Ipi2, is shown in Figure 7.3.

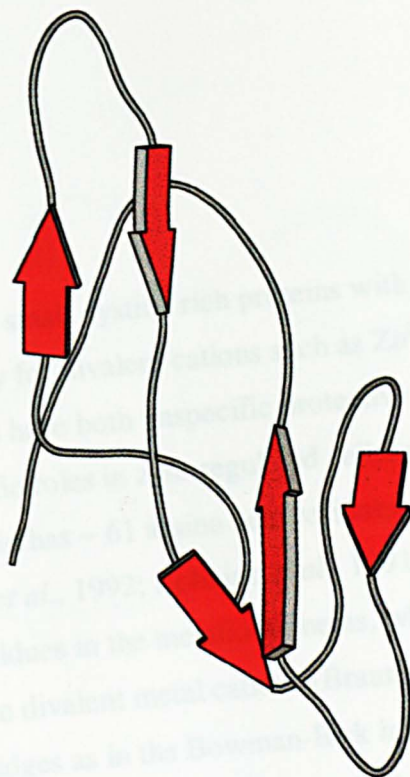


Figure 7.3 MOLSCRIPT diagram of Bowman-Birk inhibitor, Ipi2.

strong, sharp positive peak at $\sim 1680\text{ cm}^{-1}$ is suggestive of a polypeptide structure with a significant amount of disordered structure, possibly originating in PPII loop structure. Turn structure is indicated by a negative peak at $\sim 1380\text{ cm}^{-1}$. Also, a negative peak at $\sim 930\text{ cm}^{-1}$ is similar to the band found in poly(L-lysine) and may originate from left-handed helix.

The overall appearance of the ROA spectrum gives the impression of a protein with elements of ordered and disordered structure, with amino acid residues in well-defined regions of Ramachandran space, with, for example, a sharp negative ROA band at $\sim 1240\text{ cm}^{-1}$, which corresponds to the β -sheet found by UVCD (de la Sierra *et al.*, 1999).

7.5. Metallothionein

Metallothioneins are small cystine rich proteins with a molecular mass of ~ 6 kDa and have a high affinity for divalent cations such as Zn^{2+} and Cd^{2+} (Braun *et al.*, 1992). They are believed to have both unspecific protective functions against toxic metal ions as well as specific roles in zinc-regulated cellular processes (Braun *et al.*, 1992). Each metallothionein has ~ 61 amino acid residues and is devoid of any aromatic residues (Braun *et al.*, 1992; Robbins *et al.*, 1991; Schultze *et al.*, 1988). There are ~ 20 cystine residues in the metallothioneins, which are used to stabilise the tertiary fold through seven divalent metal cations (Braun *et al.*, 1992) rather than the more usual disulphide bridges as in the Bowman-Birk inhibitor. The X-ray crystal (Robbins *et al.*, 1991) and solution NMR structures (Schultze *et al.*, 1988) of rat variant metallothionein are known and are similar (Braun *et al.*, 1992). Also, the solution NMR structures of rabbit (Arseniev *et al.*, 1988) and human metallothionein (Messerle *et al.*, 1990) are known. Each variant reveals that type I and II β -turns are the dominant structures (Arseniev *et al.*, 1988; Messerle *et al.*, 1990), with human metallothionein the exception, which has $\sim 10\%$ 3_{10} -helix (Messerle *et al.*, 1990) according to NMR but not X-ray crystallography (Robbins *et al.*, 1991) (in fact 3_{10} -helix is the same as type III β -turn (Woody, 1994)). The MOLSCRIPT diagram of rat metallothionein, PDB code 4mt2, illustrates the X-ray crystal structure in Figure 7.4. Conventional Raman spectroscopy reveals that the structure of metallothionein is

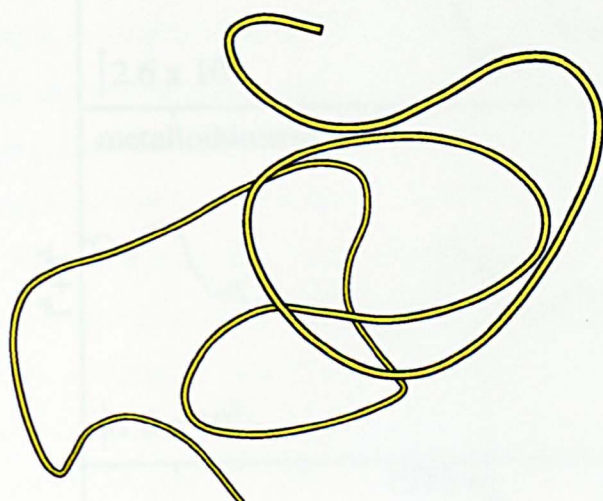


Figure 7.4 MOLSCRIPT diagram of rabbit metallothionein, 4mt2.

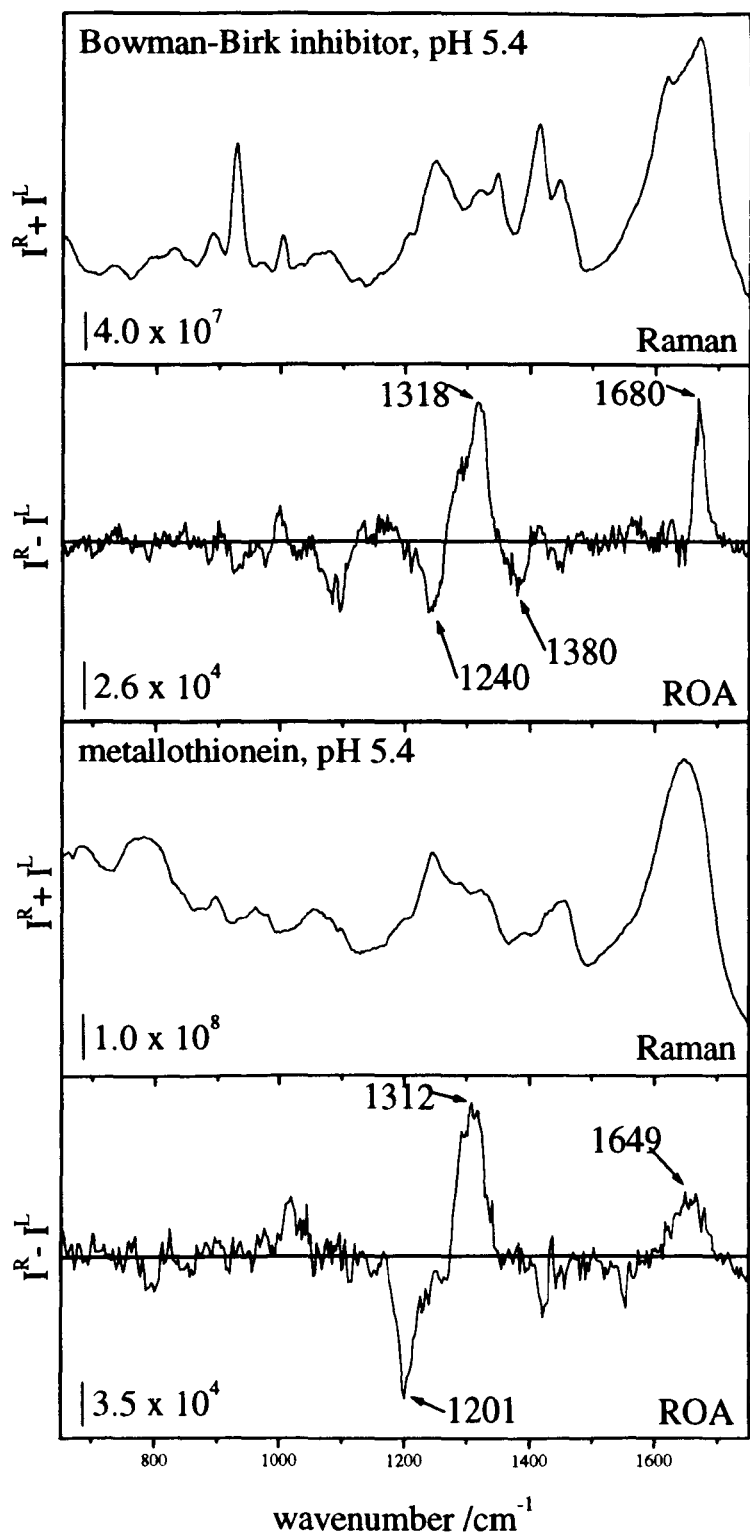


Figure 7.5 Raman and ROA spectra of Bowman-Birk inhibitor (top) and metallothionein (bottom).

consistent with a protein dominated by turn structure (Pande *et al.*, 1986).

The Raman and ROA spectra of rabbit metallothionein are presented in Figure 7.5. In some respects the ROA spectrum of metallothionein is similar to the Bowman-Birk inhibitor. The ROA spectrum of metallothionein contains sharp band structure throughout, but with fewer bands, suggesting that there is a limited range of well-defined structural types. For example, in the extended amide III region of the ROA spectrum there are only weak negative shoulders at ~ 1220 to 1260 cm^{-1} and suggests that there is little extended β -sheet structure. In addition, there is no positive band at $\sim 1340\text{ cm}^{-1}$, which is consistent with the idea that no α -helix is present. However, a broad unstructured positive peak at $\sim 1312\text{ cm}^{-1}$, may arise from a superposition of β -turn structure, as suggested by Krimm and Bandekar (1980, 1986). The amide I of the ROA spectrum is also consistent with the absence of β -sheet or α -helix. Here, a broad positive feature at ~ 1649 to 1670 cm^{-1} , consistent with disordered structure, is found. A strong, sharp negative ROA band at $\sim 1201\text{ cm}^{-1}$ is atypical for most proteins, as only a weak, negative band usually appears in this region of the ROA spectrum. Krimm has suggested that C-N stretches are found in this region and originate from type I β -turn (Krimm & Bandekar, 1986; Krimm & Bandekar, 1980; Krimm, 1987). An alternative interpretation to this is revealed by a recent ROA study on glutathione (data not shown), a tripeptide (Glu-Cys-Glu), which has a cystine residue (Creighton, 1993) that may be responsible for a sharp, negative ROA band seen at $\sim 1213\text{ cm}^{-1}$. In fact, in a previous ROA study of cysteine this band was assigned as a CH_2 wag (Gargaro *et al.*, 1993).

7.6. Discussion

In this chapter a range of protein and homopolypeptide samples with irregular folds has revealed different types of disordered solution structures. The 'unstructured' type of disorder found in phosvitin, which is similar to reduced (unfolded) proteins such as ribonuclease A and lysozyme (Wilson *et al.*, 1996a), contrasts with the 'structured' type of disorder found in the Bowman-Birk inhibitor, metallothionein and invertase which is similar to the disorder found in poly(L-lysine) and

poly(L-glutamic acid) (Wilson *et al.*, 1996b; Wilson *et al.*, 1996a). In the unstructured type of proteins, ROA data suggests a potentially wide range of (ϕ, ψ) angles for each amino acid residue, resulting in an ensemble of interconverting conformations, as expected for a random coil (Smith *et al.*, 1996). For the structured type of disorder the ROA data suggests that individual amino acid residues populate a limited range of (ϕ, ψ) angles (with small fluctuations about the mean angles). Here similarities exist between certain loop sequences in native folded proteins and proteins with irregular folds. For example, the ROA loop structure assignments in folded proteins correspond to the ROA assignments of bands in disordered polypeptides and structured irregular proteins, with the suggestion that structured disorder is comparable to the well-defined loops and turns found in globular proteins, resulting from a sequence of amino acid residues with fixed, non-repetitive (ϕ, ψ) angles (Smyth *et al.*, to be published). Unstructured and structured disorder are therefore dynamic and static, respectively.

Chapter 8

Some ROA Spectra of Proteins and Carbohydrates

In this chapter an assortment of Raman and ROA spectra of some proteins and carbohydrates is presented. Some of the proteins illustrated in this chapter do not fit with the classifications given for proteins seen in previous chapters. In addition, to demonstrate some basic ROA assignments in sugars, some carbohydrate samples are included.

8.1. Some Protein ROA spectra

8.1.1. Ribonuclease A and B

Ribonuclease A and B are digestive enzymes secreted by the pancreas and catalyse the hydrolysis of phosphodiester bonds in ribonucleic acid (RNA) (Rudd *et al.*, 1994), and both have ~ 124 amino acid residues. Ribonuclease A is nonglycosylated and has a molecular mass of ~ 13.6 kDa; whereas, ribonuclease B is glycosylated (~ 12%) and has a molecular mass of ~ 14.7 kDa (Rudd *et al.*, 1994; Joao & Dwek, 1993). There are at least five major glycans found in ribonuclease B, forming a heterogeneous mix of glycoforms, consisting of high mannose glycans with a mixture of $\alpha(1-6)$, $(1-3)$ and $(1-2)$ linked residues (Joao & Dwek, 1993; Williams *et al.*, 1987). The NMR solution structure of ribonuclease A is illustrated in Figure 8.1, and reveals a mixture of both α -helix and β -sheet (Rudd *et al.*, 1994; Joao & Dwek, 1993). X-ray crystallography studies on ribonuclease B suggested that both the glycosylated and the nonglycosylated forms of this protein have crystal structures that are equivalent, with only small structural deviations that were not considered to be statistically significant (Williams *et al.*, 1987). Although the crystal structures of ribonuclease A and B are not considered to be different, it was thought to be worthwhile to look at their solution structures to see if differences occurred in an aqueous environment. It is not inconceivable that such differences exist as, for example, CD studies on chicken

ovomucoid reveal that deglycosylation causes an increase in the solution structure of its β -sheet content, relative to its glycosylated variant (Watanabe *et al.*, 1981).

The Raman and ROA spectra of ribonuclease A and B are presented in Figure 8.2. The ROA spectrum of ribonuclease B is rather noisy, but still of a sufficient quality to make comparisons. The backbone region of the ROA spectrum of ribonuclease A is well resolved and clearly shows a positive band at $\sim 831\text{ cm}^{-1}$ to 909 cm^{-1} and is indicative of α -helix. A similar feature exists in the ROA spectrum of ribonuclease B, but is rather less well-defined. The negative-positive couplet centred at $\sim 1267\text{ cm}^{-1}$ is present in the ROA spectra of both samples and is thought to be a signature of α -helix (Wilson, 1996). Also, in the amide I, a negative peak at $\sim 1649\text{ cm}^{-1}$ is consistent with the presence of α -helix (Wilson, 1996). The β -sheet structure of these proteins is indicated by a positive ROA band at $\sim 1673\text{ cm}^{-1}$ (Wilson, 1996). In the extended amide III region of the ROA spectrum a negative band at $\sim 1245\text{ cm}^{-1}$ in both samples suggests the presence of ordered β -sheet structure (Barron *et al.*, in press; Smyth *et al.*, to be published). A positive broad band at $\sim 1315\text{ cm}^{-1}$ to 1320 cm^{-1} in both ROA spectra also suggests that the PPII type structure is present. Also, a positive ROA band at $\sim 1001\text{ cm}^{-1}$ to 1048 cm^{-1} in ribonuclease A has been tentatively assigned as originating in β -sheet composed of right-twist β -strands (Wilson, 1996). As before, this region is less well resolved in the ROA spectrum of ribonuclease B. Due to the relatively low carbohydrate content of ribonuclease B, there are no discernible ROA bands originating from glycan vibrations.

8.1.2. Insulin

In 1935 Dorothy Hodgkin crystallised insulin (Perutz, 1992). Being such a small molecule, with a molecular mass of $\sim 6\text{ kDa}$, it was thought that the X-ray crystal structure would be easy to solve (Perutz, 1992). In fact, it took Hodgkin another 34 years to achieve this (Perutz, 1992). Insulin consists of two polypeptide chains, with one chain containing 21 (the A-chain) and the other 30 (the B-chain) amino acid residues (Perutz, 1992; Lesk, 1991). The structure of insulin is illustrated by the MOLSCRIPT diagram in Figure 8.3. Two disulphide bonds (A7-B7 and A20-B19) link both chains with a third intramolecular disulphide bond in the A-

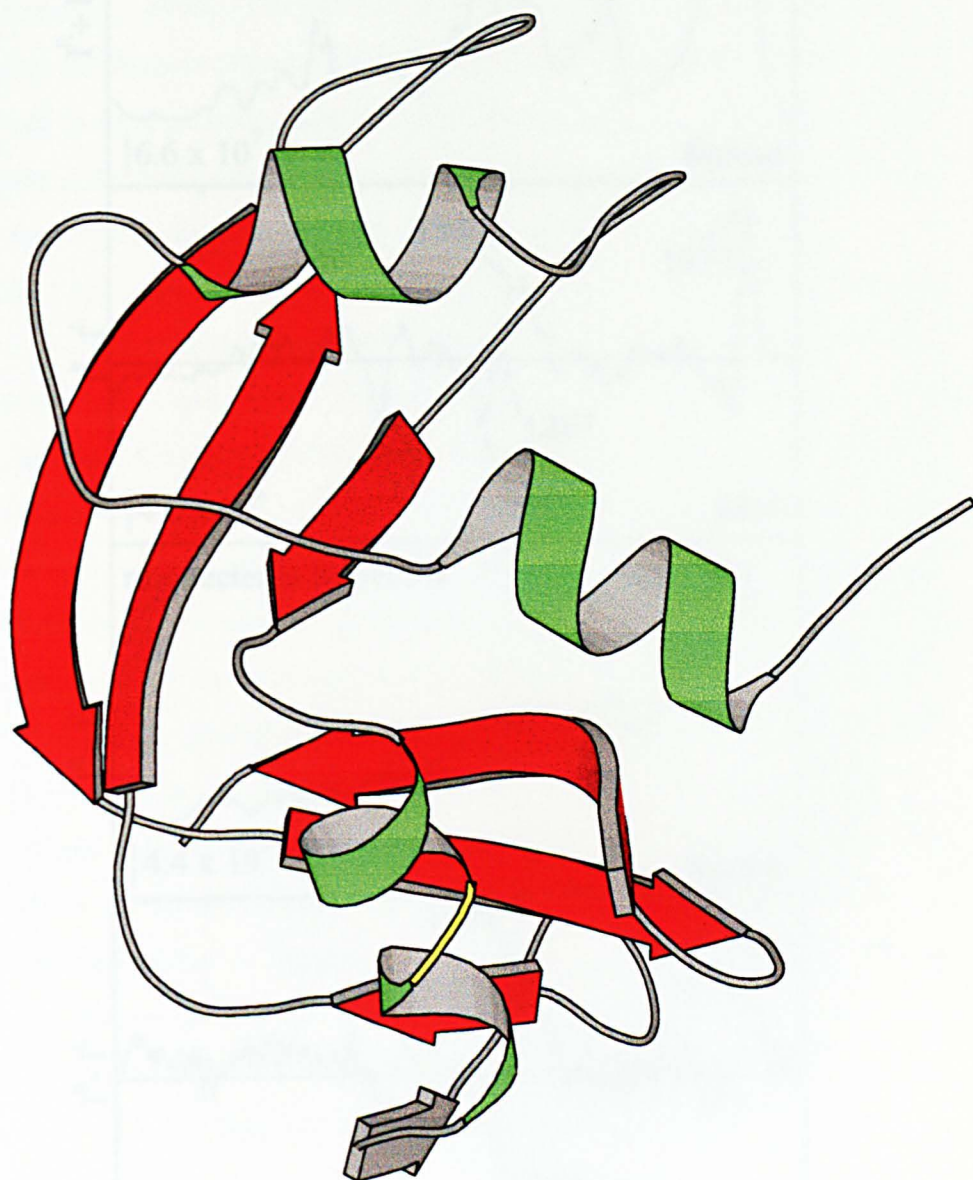


Figure 8.1 MOLSCRIPT diagram of the solution structure of ribonuclease A, 1rat.

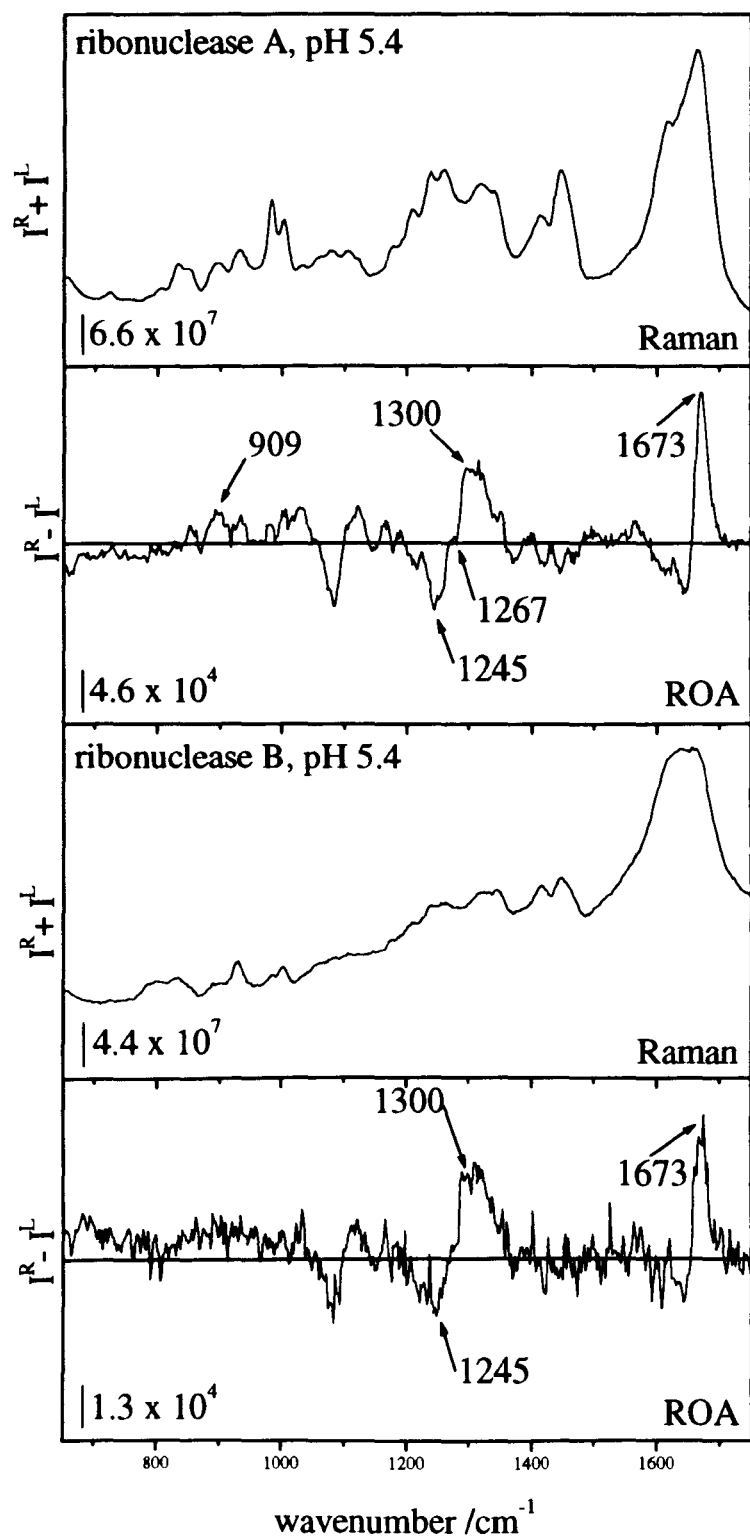


Figure 8.2 Raman and ROA spectra of ribonuclease A (top) and ribonuclease B (bottom).

chain (A6-A11) (Perutz, 1992; Lesk, 1991). In solution insulin self-associates to form a dimer, and a hexamer (formed from three dimers) (Perutz, 1992; Lesk, 1991). The Higher-order oligomeric forms of insulin are induced by increasing the pH and the concentration of the sample, as well as coordination to Zn^{2+} (Perutz, 1992; Lesk, 1991; Jørgensen *et al.*, 1996). Dimerisation is facilitated by the highly flexible C-terminal end of the B-chain (Perutz, 1992; Lesk, 1991), which associates with a second B-chain to form an antiparallel hydrogen bonded β -sheet. The *in vivo* active form of insulin is the monomer (Perutz, 1992), which has serious consequences for sufferers of type I diabetes, with synthetic forms of insulin manufactured as a hexamer, which has to be broken down to its monomeric form before it can be used, resulting in a delayed uptake of \sim two hours (Perutz, 1992). Engineered (mutated) forms of insulin exist (Perutz, 1992; Brange *et al.*, 1988) with a faster uptake of \sim one hour. If an aspartate residue is substituted in place of proline (B28), electrostatic repulsion between the boundary of the monomers raises the dissociation constant of the hexamer into monomers (Perutz, 1992; Brange *et al.*, 1988). Dissociation of the hexamer into a monomer is required as the face of the insulin molecule that binds to the insulin receptor is the same as that which forms the contact between the dimers (Perutz, 1992). X-ray crystallography reveals that the structure of hexameric insulin contains \sim 41.2% α -helix and 5.9% 3_{10} -helix (Badger *et al.*, 1991). Also, NMR (Perutz, 1992) and X-ray studies (Bi *et al.*, 1984) indicate that the hexameric and the monomeric forms are identical, within experimental error. The Raman and ROA spectra of insulin are shown in Figure 8.4. Dynamic light scattering studies (data not shown) suggest that the dimeric and monomeric forms of insulin were present during this ROA study at pH 2.2 and 0.8, respectively. In the ROA spectrum of dimeric insulin a positive peak at $\sim 900\text{ cm}^{-1}$ to 974 cm^{-1} confirms the presence of α -helix. In the extended amide III region of the ROA a positive band at $\sim 1308\text{ cm}^{-1}$ to 1310 cm^{-1} is in a region contiguous with α -helix (Wilson, 1996). However, α -helix is normally shifted down by $\sim 8\text{ cm}^{-1}$ to 10 cm^{-1} and is found at $\sim 1300\text{ cm}^{-1}$ (Wilson, 1996). In a Raman study by Krimm it was suggested that this region is also likely to contain vibrations from β -turns (Krimm & Bandekar, 1980), which could superpose with any α -helix vibrations in this region of the ROA spectrum. A strong sharp negative ROA band at $\sim 1244\text{ cm}^{-1}$ suggests the presence of β -strand, which possibly originates from the C-terminal β -strand of the B-chain. In the dimeric form of insulin this β -strand is

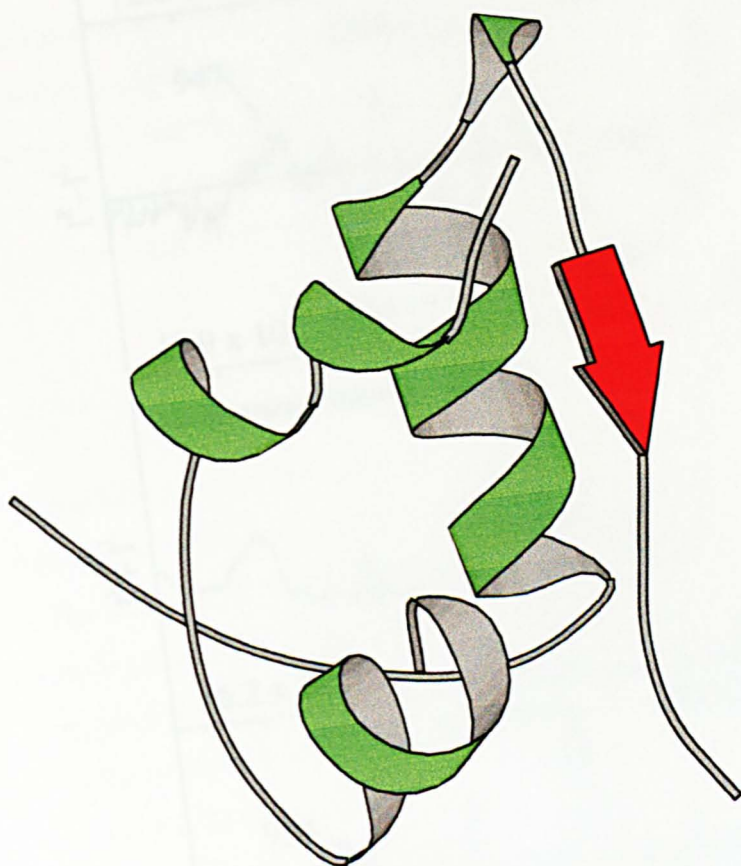


Figure 8.3 MOLSCRIPT diagram of insulin, 9ins.

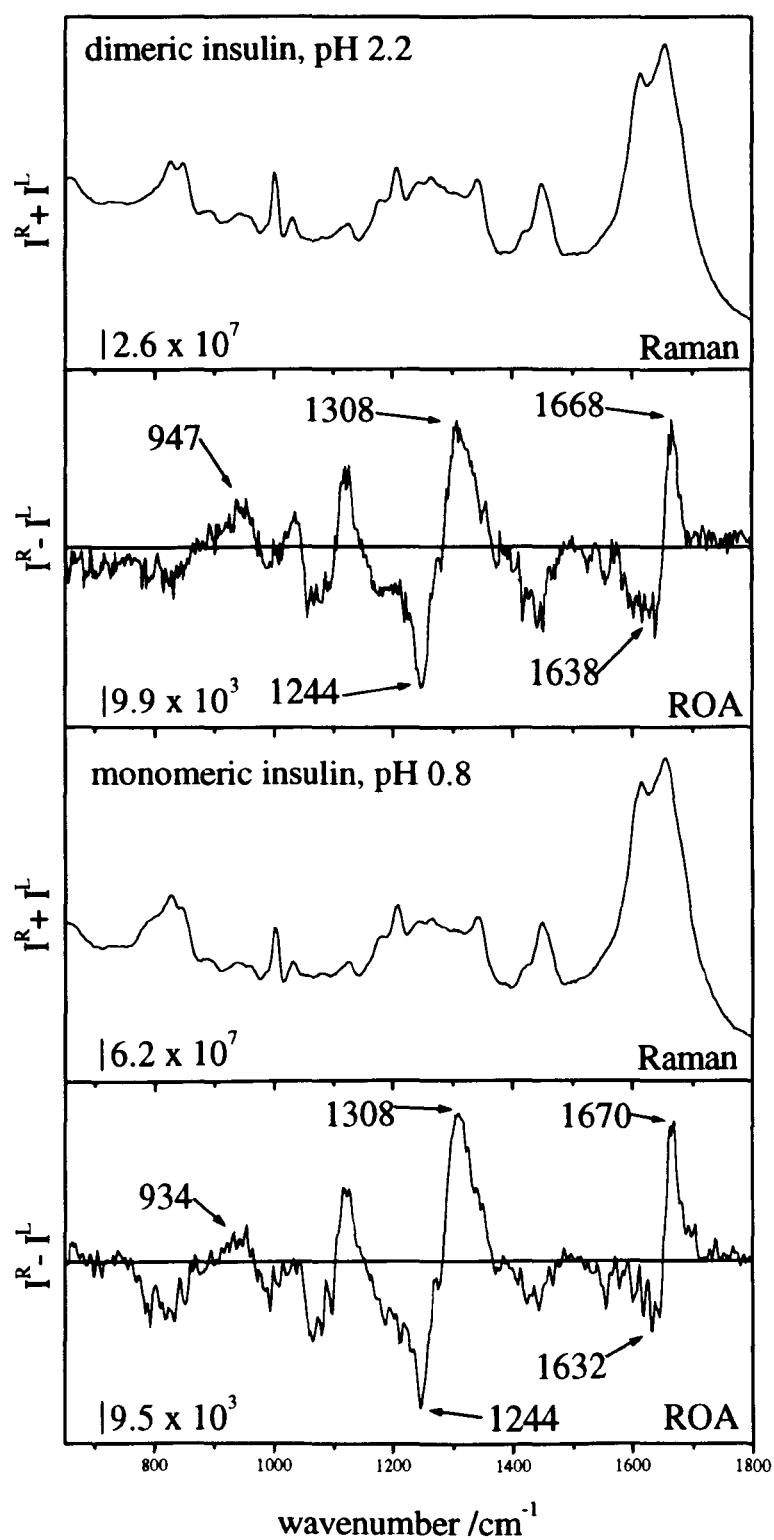


Figure 8.4 Raman and ROA spectra of insulin in H_2O at pH 2.2 (top) and pH 0.8 (bottom).

involved in a double strand β -sheet with its two outer edges exposed to the solvent (Perutz, 1992). The amide I region of the ROA spectrum is difficult to interpret. A negative-positive couplet, negative at $\sim 1638\text{ cm}^{-1}$ and positive at $\sim 1668\text{ cm}^{-1}$, is suggestive of a protein structure with α -helix, β -turn structure and β -strand (Wilson, 1996). Raman studies by Krimm suggest that turn structure is commonly found at $\sim 1640\text{ cm}^{-1}$ for type I β -turn, whereas, α -helix is $\sim 10\text{ cm}^{-1}$ higher in the Raman spectrum (Krimm & Bandekar, 1980; Krimm & Bandekar, 1986). The positive 1668 cm^{-1} ROA band is near to a region normally assigned as β -sheet. The monomeric form of insulin was also studied to see if structural differences existed between the dimer and the monomer. It is readily seen that both sets of spectra are similar. This is an important fact in light of the earlier discussion, as ROA could be used to observe structural rearrangements in engineered forms of this hormone.

8.1.3. Ovotransferrin (Conalbumin)

Diferric hen ovotransferrin, also known as type I conalbumin, is a monomeric glycoprotein with a molecular mass of $\sim 80\text{ kDa}$ (Schneider *et al.*, 1984; Kurokawa *et al.*, 1995). The structure of diferric ovotransferrin is illustrated in Figure 8.5. In this study an iron-depleted form of the protein was used (illustrated in Figure 8.6), as the diferric form has a bright pink colour, which would cause the sample to burn up in the laser beam. Due to its avidity for iron ions (Schneider *et al.*, 1984; Kurokawa *et al.*, 1995) ovotransferrin is an excellent antibacterial agent, and acts by sequestering iron ions from egg white fluids, as iron ions are essential for bacterial growth. The molecule is divided into two homologous lobes (N- and C-lobes), each containing a single iron-binding site (Kurokawa *et al.*, 1995). Each lobe is further subdivided into two dissimilar domains (N1 and N2 for the N-terminal lobe and C1 and C2 for the C-terminal lobe), with iron bound between the interdomain cleft together with a synergistic ion such as bicarbonate (Kurokawa *et al.*, 1995). The ROA spectra of ovotransferrin in H_2O and D_2O are shown in Figure 8.7. It is unfortunate that only the iron-depleted form of ovotransferrin is available for study as large-scale conformational changes are known to take place (Kurokawa *et al.*, 1995). When iron is released the interdomain clefts are opened and made ready to accept a new guest ion (see Figures 8.6 and 8.7). The ROA spectrum of ovotransferrin in H_2O is unusual, as

the polypeptide backbone region at $\sim 850\text{ cm}^{-1}$ to 950 cm^{-1} is rather weak for a protein with a high α -helix content of $\sim 28\%$ versus a β -sheet content of $\sim 18\%$ (Kurokawa *et al.*, 1995). A prominent positive ROA band in the extended amide III at $\sim 1316\text{ cm}^{-1}$ is too high to be associated with α -helix and is more indicative of PPII loop structure (Barron *et al.*, in press). However, a moderately strong positive band at $\sim 1339\text{ cm}^{-1}$ suggests that some highly hydrated α -helix is present. In the D_2O ROA spectrum the 1339 cm^{-1} band is significantly diminished, demonstrating its assignment as solvent exposed helical structure. An interesting feature in this region is the lack of a negative ROA band at $\sim 1360\text{ cm}^{-1}$, which is normally assigned as β -turn structure (Wilson, 1996). X-ray crystallography studies show that ~ 35 to 40% of the structure of diferric ovotransferrin (Kurokawa *et al.*, 1995; pdb, 1ovt) is composed of turn structure. Such anomalies in the α -helix and β -turn regions of the ROA spectrum may be attributable to the fact that the iron-depleted form is studied here. A broad negative ROA band centred at $\sim 1245\text{ cm}^{-1}$ with a weak shoulder at $\sim 1222\text{ cm}^{-1}$ is indicative of β -sheet and solvated β -sheet, respectively. In the D_2O ROA spectrum two sharp negative ROA bands appear at $\sim 1216\text{ cm}^{-1}$ and 1265 cm^{-1} , with the 1245 cm^{-1} band significantly reduced. This suggests that the β -sheets of the iron-depleted ovotransferrin are rather ill-defined and highly solvated. In the amide I of the ROA spectrum a strong positive band at $\sim 1672\text{ cm}^{-1}$ suggests that in solution β -sheet structure has a significant presence (Wilson, 1996; Barron *et al.*, in press). Overall, the ROA spectrum of iron-depleted ovotransferrin is intriguing as it suggests that the solution state of this biomolecule has a high degree of disorder. The reasons for this are unclear. The iron-bound form may exhibit a larger degree of order in keeping with a more closed structure as revealed by X-ray crystallography (see earlier remark). Alternatively, solvation effects may be perturbing the solution structure from its solid-state conformation.



Figure 8.5 MOLSCRIPT diagram of diferric ovotransferrin, 1ovt.

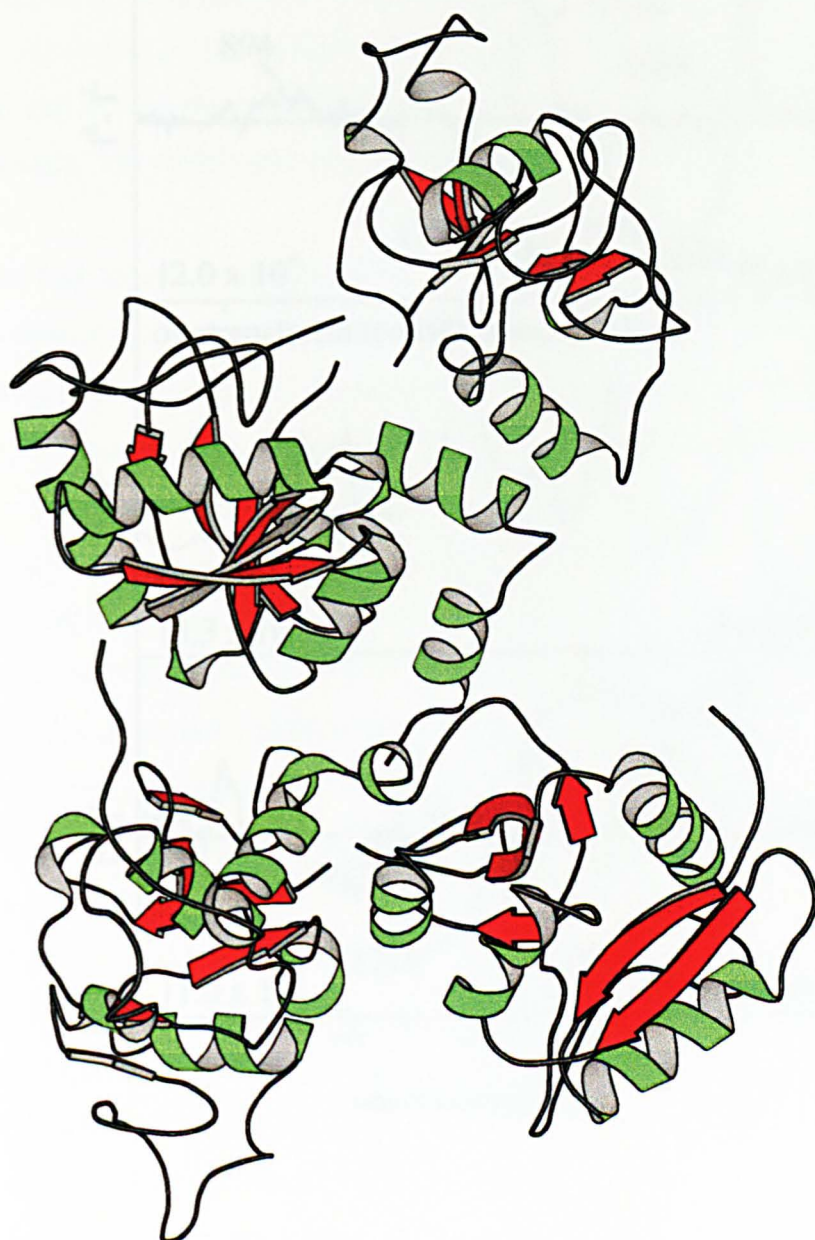


Figure 8.6 MOLSCRIPT diagram of apo-ovotransferrin, 1aov.

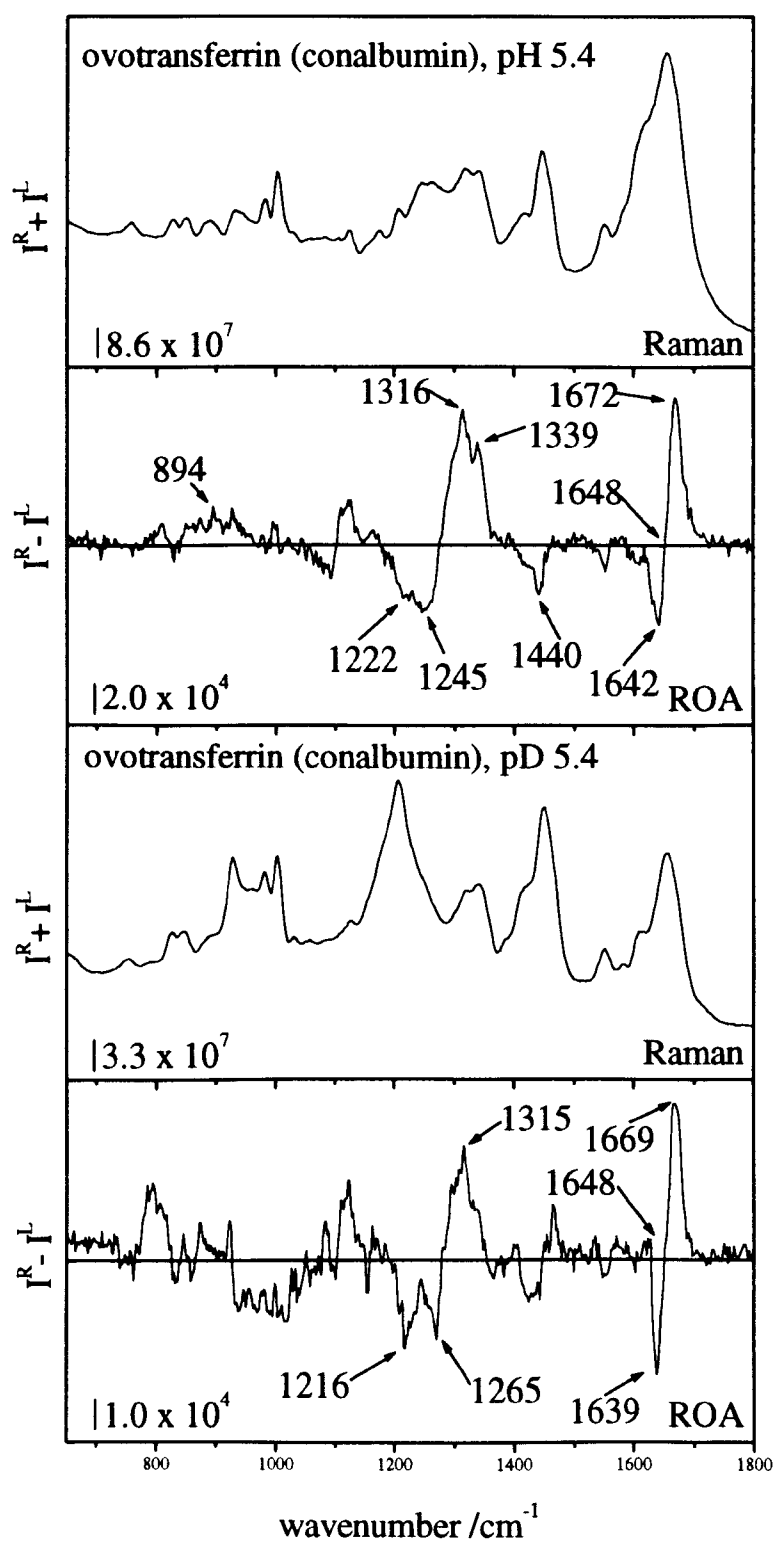


Figure 8.7 Raman and ROA spectra of ovotransferrin in H_2O (top) and D_2O (bottom).

8.1.4. Bovine serum albumin (BSA), Human serum albumin (HSA) and Glycosylated (?) α_1 -Antiproteinase

The serum albumins (bovine and human) are principal component of blood and are responsible for the maintenance of blood pH (Carter *et al.*, 1989) and the facilitated transfer of ligands across organ-circulatory interfaces (He & Carter, 1992). They are large proteins with ~ 585 amino acid residues (Carter *et al.*, 1989; He & Carter, 1992) and a molecular mass of ~ 65 kDa, and contain 17 disulphide bonds. It is assumed that the structure of BSA is similar to HSA which has a high helix content of ~ 67% (He & Carter, 1992). HSA is composed of three homologous domains assembled into a heart-shaped structure (He & Carter, 1992) with six subdomains, two in each domain. The structure of HSA is illustrated by the MOLSCRIPT diagram in Figure 8.8.

The Raman and ROA spectra of BSA in H₂O and HSA with glycosylated (?) α_1 -antiproteinase are shown in Figures 8.9 and 8.10, respectively. The Raman and ROA spectra of glycosylated (?) α_1 -antiproteinase is included to provide a cautionary note on sample purity. Also, there are no D₂O Raman or ROA spectra of BSA, HSA or glycosylated (?) α_1 -antiproteinase, due to instrument problems at the time of obtaining these spectra. Recent studies on HSA have managed to overcome this problem. If the ROA spectra of BSA and HSA (both in H₂O) are compared it is readily seen that both spectra are remarkably similar, which confirms the assumption on their structural similarity made at the start of this section. Three α -helix signatures appear in both spectra. A positive ROA band in the skeletal stretch region at ~ 850 cm⁻¹ to 960 cm⁻¹, the broad positive ROA band at ~ 1300 cm⁻¹ which in fact contains two peaks at ~ 1297 cm⁻¹ and at ~ 1308 cm⁻¹, and a strong positive ROA band at ~ 1340 cm⁻¹ in HSA and at ~ 1342 cm⁻¹ in BSA. In a previous D₂O study (Wilson, 1996) of BSA the 1342 cm⁻¹ band disappeared and indicates that this band is sensitive to solvation. A similar result is obtained for HSA (data not shown). Earlier ROA work on HSA (Teraoka *et al.*, 1998), at a pH of ~ 3.4 to induce the molten globule F-state, revealed that ~ 40% of the intensity of the 1340 cm⁻¹ band is lost, and is ascribed to a loss of rigidity of some of the α -helix structure, which is thought to be highly hydrated (Teraoka *et al.*, 1998). A negative ROA band, which is rather broad



Figure 8.8 MOLSCRIPT diagram of HSA, 1ao6.

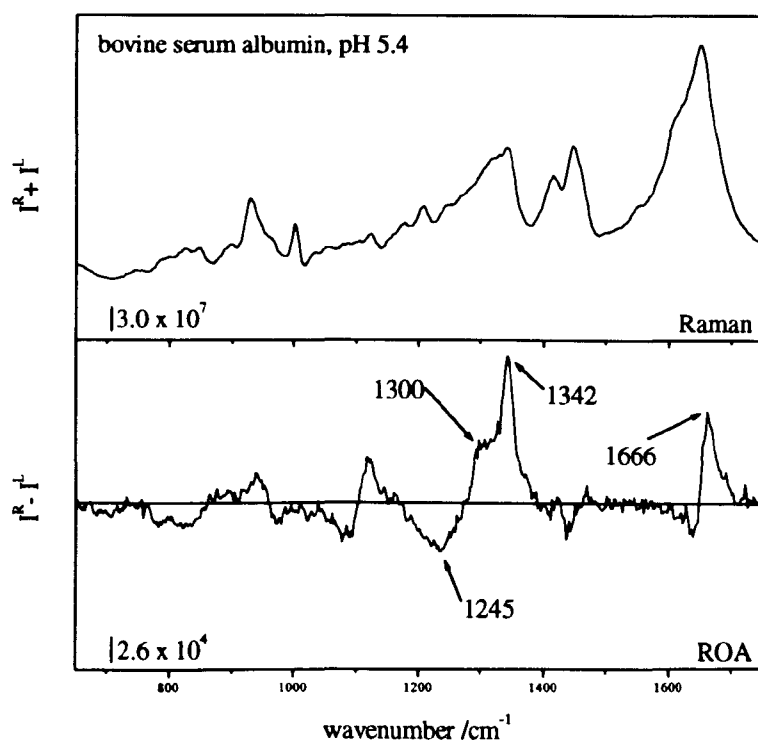


Figure 8.9 Raman and ROA spectra of BSA in H_2O .

in BSA but better defined in HSA, is also seen in both spectra at $\sim 1245\text{ cm}^{-1}$ and suggests that some β -structure is present in both samples. Also, the amide I couplet of HSA shows a negative ROA band at $\sim 1640\text{ cm}^{-1}$ and a positive band at $\sim 1666\text{ cm}^{-1}$ and is similar to those seen in lysozyme and α -lactalbumin, which also originate mainly in α -helix (Wilson, 1996; Barron *et al.*, in press). Finally, the ROA and Raman spectra of glycosylated (?) α_1 -antiproteinase is seen to be (although rather noisy) very similar to the serum albumin spectra. When the ROA spectrum of this sample was first run, it suggested that deglycosylation (there are three sites of glycosylation (Asn 46, 83 and 247 (Huber & Carrell, 1989)) of the polypeptide backbone, resulted in a substantial loss of α -helix structure in the non-glycosylated recombinant form of α_1 -antiproteinase (see chapter 6 for the ROA spectrum). However, it now seem likely that the resemblance of glycosylated (?) α_1 -antiproteinase to the serum albumins is more than coincidental. Upon making telephone enquiries to Sigma, the purity of this sample seems doubtful, as a significant proportion of this sample (as much as 50%) is contaminated with HSA. Earlier preparations of α_1 -antiproteinase had a higher sample purity of $\sim 90\%$ but are no longer available.

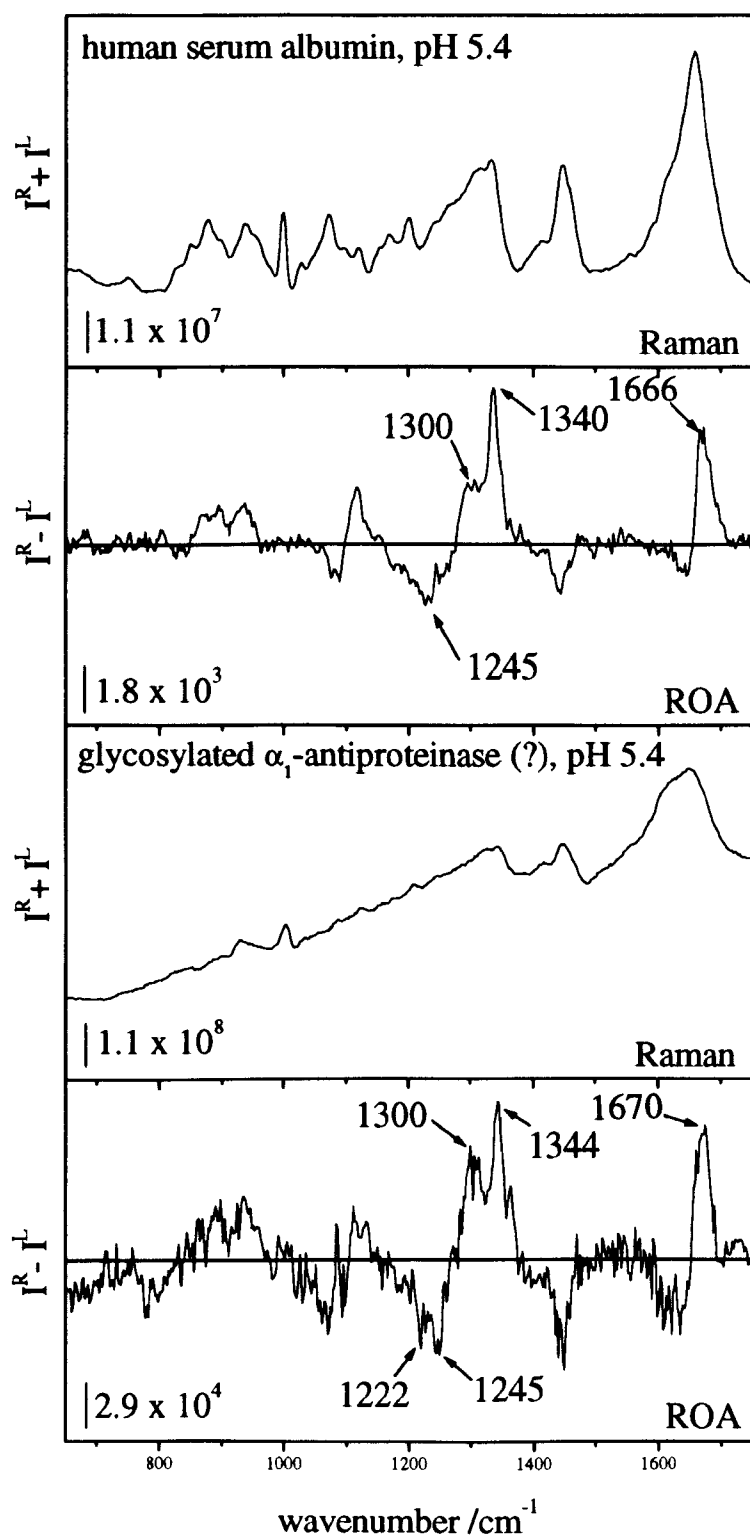


Figure 8.10 Raman and ROA spectra of HSA (top) and glycosylated (?) α_1 -antiproteinase.

8.1.5. Aldolase

The glycolytic enzyme aldolase has an $(\alpha/\beta)_8$ structure with a triose phosphate isomerase (TIM) barrel-like form (Lesk, 1991; Sygusch *et al.*, 1987), and is chiefly responsible for the transformation of dihydroxyacetone phosphate to D-glyceraldehyde 3-phosphate (Belasco & Knowels, 1980). It contains 8 parallel α/β subunits, with the strands forming a β -sheet closed into a cylinder or barrel, and has 8 α -helices packed against the outside of the sheet (Lesk, 1991). An active site is located at the end of the barrel that corresponds to the C-termini of the β -sheet. Numerous TIM barrels exist and catalyse a variety of different reactions (Lesk, 1991). Some general properties of the TIM barrel structure are as follows: each of the 8 β -strands that make up the barrel are in a parallel arrangement with each strand tipped by approximately 36 degrees to the barrel axis (Lesk, 1991; Chou & Carlacci, 1991). The helices are approximately parallel to the strands, with the chain always advancing in an anticlockwise sense (Lesk, 1991; Chou & Carlacci, 1991) and, when viewed locally from the C-termini, the chain proceeds 'up' a strand and 'down' a helix. Also, the regions of the polypeptide chain between different secondary structure elements are soft except for the 'helix-loop' regions (Kobayashi *et al.*, 1997), with the ' β -barrel interior' found to be strikingly soft. The 'hard' helix-loop regions are thought to be more rigid than ice and the β -sheet interior is as 'soft' as water (Kobayashi *et al.*, 1997). The TIM barrel motif is one of the most common which demonstrates its great utility as an enzyme (Lesk, 1991). The α -helix and β -strand content of aldolase is ~40% for both structural elements (Blom & Sygusch, 1997). The remainder of the structure consists of 3_{10} -helix and β -turn structure (Blom & Sygusch, 1997). A MOLSCRIPT diagram of aldolase illustrates its structure in Figure 8.11.



Figure 8.11 MOLSCRIPT diagram of aldolase, Iado.

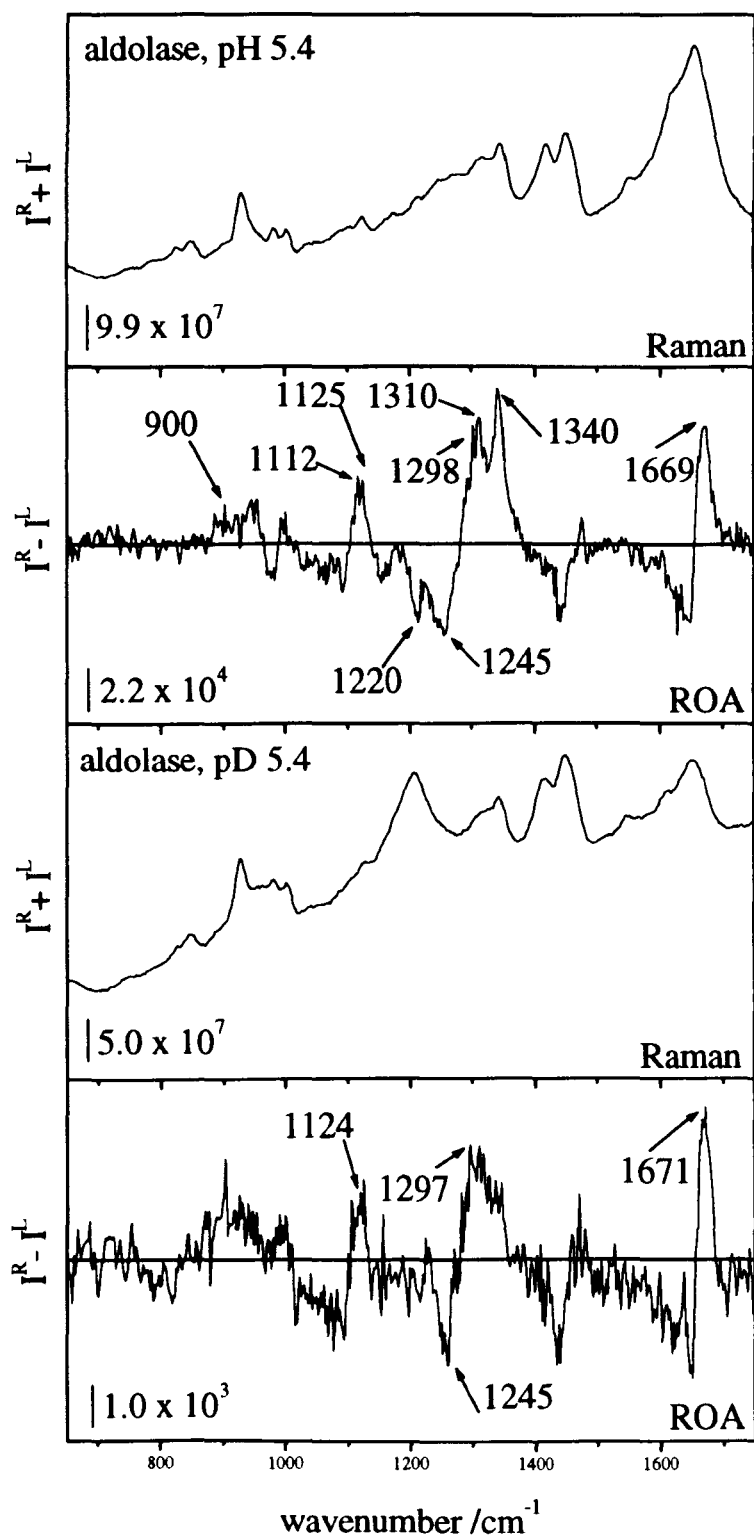


Figure 8.12 Raman and ROA spectra of aldolase in H_2O (top) and D_2O (bottom).

The Raman and ROA spectra of aldolase in H₂O and D₂O are shown in Figure 8.12. The ROA spectrum of aldolase is dominated by a strong, positive peak at $\sim 1340\text{ cm}^{-1}$ and is assigned as originating in the α -helices that form a rosetta-like crown around the central β -sheet. Looking at the MOLSCRIPT diagram of aldolase, it can be seen that the helices have their inside edges directed towards the core of the molecule, which is possibly non-hydrated, whereas the outside edge of the helices is directed towards the solvent. Consequently, it is suggested that the 1340 cm^{-1} ROA band is generated by highly hydrated α -helices (Barron *et al.*, in press). When aldolase is deuterated this band disappears, revealing the highly solvated nature of the structural elements that generate this band. Two weak, positive ROA bands at $\sim 870\text{ cm}^{-1}$ and $\sim 940\text{ cm}^{-1}$ in the backbone skeletal stretch region confirm the presence of α -helix. A positive ROA band centred at $\sim 1122\text{ cm}^{-1}$ is split into a doublet at $\sim 1112\text{ cm}^{-1}$ and 1125 cm^{-1} and is typical of α/β proteins. Two positive ROA bands are seen at 1298 cm^{-1} and 1310 cm^{-1} , probably originate in α -helix modes, and not from turn structure as in insulin (see earlier in this chapter), due to the significant amount of α -helix found in aldolase. The D₂O ROA spectrum of aldolase would seem to confirm this as these bands are unaffected by deuteration. This behaviour suggests that the $1300\text{ to }1309\text{ cm}^{-1}$ α -helix mode is less sensitive to solvation when compared to the 1340 cm^{-1} band.

A sharp negative ROA band at $\sim 1245\text{ cm}^{-1}$ is associated with the β -strands of the molecule. Previously, bands in this region were thought to originate in PPII loop structure (Barron *et al.*, in press), however, the absence of any PPII in aldolase would seem to rule this out. A slightly weaker negative ROA band at $\sim 1220\text{ cm}^{-1}$ is also associated with β -sheet (see earlier remarks in this chapter), with hydrated β -sheet suggested as one possibility. The D₂O ROA spectrum of aldolase reveals that the negative 1245 cm^{-1} ROA band is relatively intact after deuteration, but that the negative 1220 cm^{-1} ROA band is lost. It is possible that the 1245 cm^{-1} ROA band is generated by β -sheet strands buried deep within the β -sheet core of aldolase, which is known to be highly hydrophobic (Kobayashi *et al.*, 1997) and, that the 1220 cm^{-1} ROA band is generated by end residues of β -strands near the solvent/protein interface. An interesting feature of aldolase is the absence of a negative ROA band at $\sim 1360\text{ cm}^{-1}$ which is, when present, assigned as β -turn and suggests that little turn structure is present and is indicative of the fact that the barrel structure of aldolase is composed

of parallel β -strands connected by interleaved α -helices. This feature will be discussed in the next section.

8.1.6 Concanavalin A

Although concanavalin A (conA) is properly classed as a β -sheet protein (see chapter 5) it is better placed in this chapter to give a direct comparison against aldolase. Concanavalin A is a member of the lectin family and is a carbohydrate binding protein consisting of four identical, individual monomers each of molecular mass ~ 25 kDa, at physiological pH (Edmund *et al.*, 1971; Wang *et al.*, 1971). Each monomer has an individual carbohydrate binding site meaning that four saccharides can be bound by the tetramer (So & Goldstein, 1968; Yarin *et al.*, 1968). The structure of conA is formed from a 'jelly roll' of eight β -strands in an antiparallel β -sheet arrangement (Branden & Tooze, 1991; Becker *et al.*, 1975; Reeke *et al.*, 1975; Weisgerber & Helliwell, 1993). The jelly roll structure is divided into two large β -sheets with few hydrogen bonds between strands that belong to different sheets (Branden & Tooze, 1991). Topologically the jelly roll can be viewed as two 'Greek key' motifs (Branden & Tooze, 1991). The folding of conA is mediated by a long continuous antiparallel hairpin that twists and curls to form a barrel structure with loops that protrude from the 'top' and 'bottom' of the molecule (Branden & Tooze, 1991; Becker *et al.*, 1975; Reeke *et al.*, 1975; Weisgerber & Helliwell, 1993) (see MOLSCRIPT diagram, Figure 8.13).

The Raman and ROA spectra of conA in H_2O is shown in Figure 8.14. The backbone skeletal stretch region of the ROA spectrum of conA in H_2O is unusual as prominent bands appear at $\sim 900\text{ cm}^{-1}$ to $\sim 1000\text{ cm}^{-1}$ indicating that a substantial amount of α -helix is present, which is at odds with X-ray crystallography. A strong, sharp negative ROA band at $\sim 1237\text{ cm}^{-1}$ is indicative of β -strands and a weaker negative ROA band at $\sim 1200\text{ cm}^{-1}$ is also seen. A strong, positive ROA band is observed at $\sim 1316\text{ cm}^{-1}$ and is now thought to originate in PPII loop structure (Barron *et al.*, in press; Smyth *et al.*, to be published). A broad ill defined positive shoulder from $\sim 1288\text{ cm}^{-1}$ to 1308 cm^{-1} is suggestive of β -turn structure. In addition, a strong,

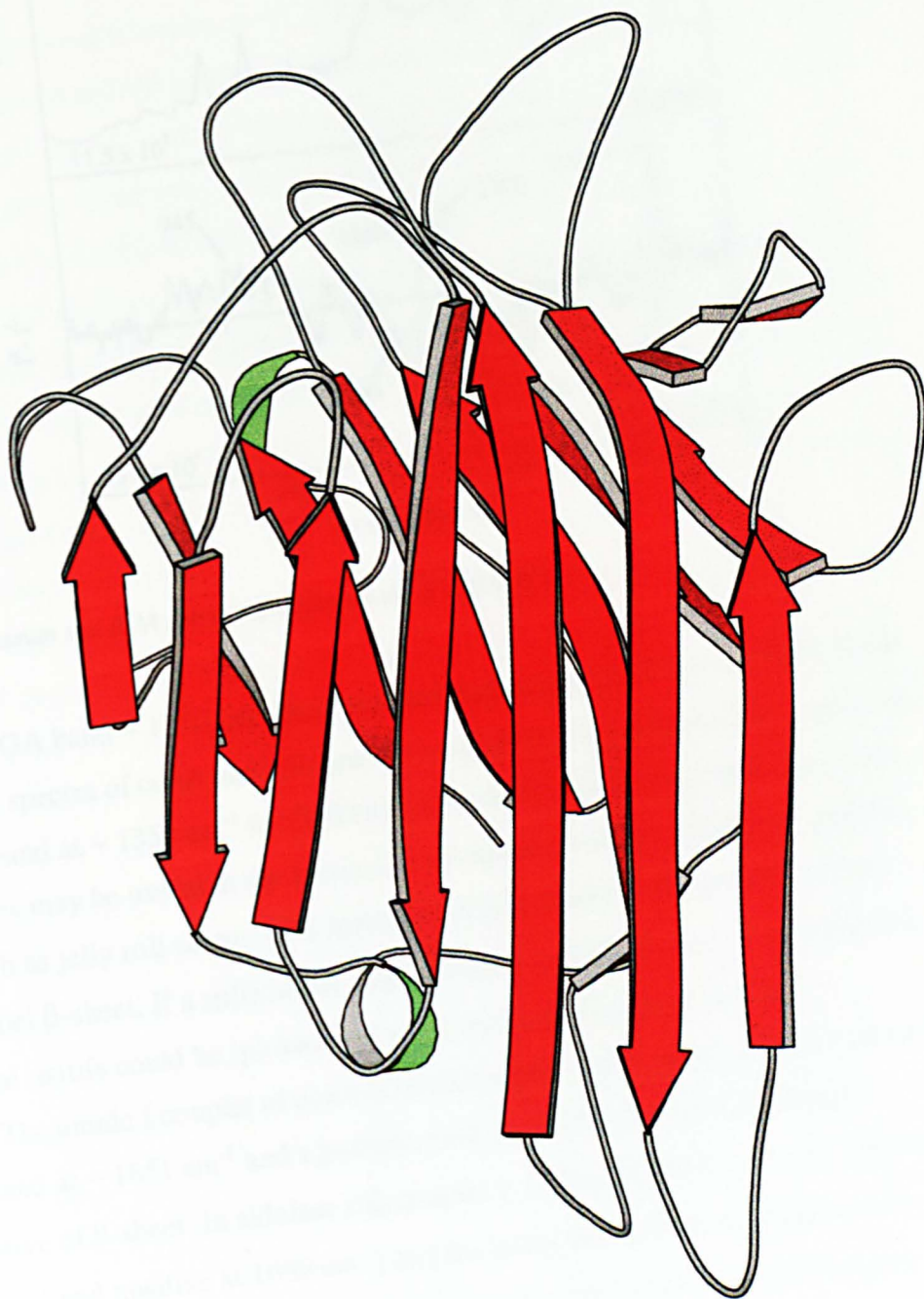


Figure 8.13 MOLSCRIPT diagram of concanavalin A, 2cna.

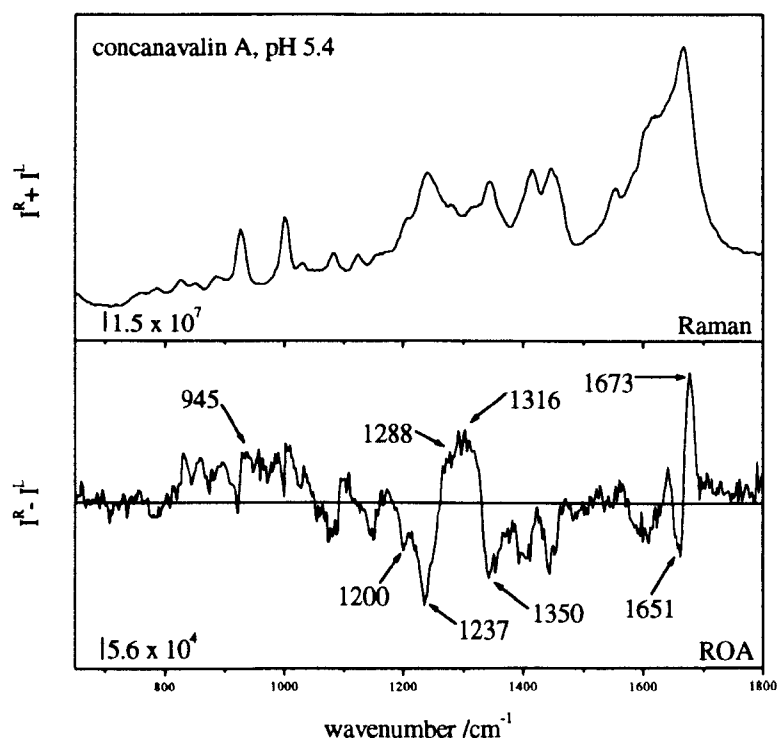


Figure 8.14 Raman and ROA spectra of concanavalin A and in H_2O .

negative ROA band $\sim 1350 \text{ cm}^{-1}$ has been assigned as β -turn (Wilson, 1996). If the H_2O ROA spectra of conA and aldolase are compared, it is apparent that there is no negative band at $\sim 1350 \text{ cm}^{-1}$ to 1360 cm^{-1} in the ROA spectrum of aldolase. Such differences may be useful in the future as a means of identifying different protein folds such as jelly roll *versus* TIM barrels, that is in determining parallel *versus* antiparallel β -sheet. If a sufficiently large database of ROA spectra is constructed, structural motifs could be ‘picked out’ and used to identify protein folds.

The amide I couplet of conA is uncharacteristically sharp and has a negative ROA band at $\sim 1651 \text{ cm}^{-1}$ and a positive peak at $\sim 1673 \text{ cm}^{-1}$ and is strongly suggestive of β -sheet. In aldolase this couplet is shifted down by $\sim 4 \text{ cm}^{-1}$ (negative at 1645 cm^{-1} and positive at 1669 cm^{-1}) and indicates that both β -sheet and α -helix structures are present. Again, such differences may be useful in discriminating between different types of β -structural motifs.

8.1.7. Type I Dehydroquinase

Type I dehydroquinase is a TIM barrel protein with a mass of ~ 28 kDa and ~ 240 amino acid residues (Gourley *et al.*, 1999; Harris *et al.*, 1996). A type II variant of dehydroquinase exists (Gourley *et al.*, 1999; Harris *et al.*, 1996) with an α - β - α sandwich structure. Although unrelated at the sequence level (Kleanthous *et al.*, 1992) and with different physical and biochemical properties (Harris *et al.*, 1993; Harris *et al.*, 1996), both of these enzymes catalyse the same dehydration reaction of 3-dehydroquininate **1** to 3-dehydroshikimate **2** (Bentley, 1990). It is normally assumed that individual enzymes are optimised to catalyse reactions by particular (specific) pathways (Harris *et al.*, 1993). Here Type I and II dehydroquinase catalyse the same reaction as a consequence of convergent evolution, but with different mechanistic details (Kleanthous *et al.*, 1992). Clearly, these enzymes are of interest, as an understanding of their mechanistic properties could be used to produce enzymes with enhanced catalytic properties.

The Raman and ROA spectrum of type I dehydroquinase in H_2O is shown in Figure 8.15. Note that the ROA spectrum of dehydroquinase is rather noisy and is displayed above 900 cm^{-1} due to baseline problems. Nevertheless, it was felt that it was useful to study this sample to see if comparisons could be made with aldolase. In a similar manner to aldolase a negative/positive couplet centred at $\sim 1100\text{ cm}^{-1}$ appears, but is more intense. In the extended amide III region of the ROA spectrum a number of negative bands with large noise spikes at $\sim 1220\text{ cm}^{-1}$, $\sim 1236\text{ cm}^{-1}$ and $\sim 1252\text{ cm}^{-1}$ suggest the presence of β -strands. However, in the corresponding aldolase spectrum this region is better resolved, and it is difficult to make an exact comparison with dehydroquinase. At slightly higher wavenumbers two positive ROA bands at $\sim 1302\text{ cm}^{-1}$ and $\sim 1341\text{ cm}^{-1}$ are better resolved and are quite similar in their appearance to the corresponding ROA bands in aldolase; the only slight difference being that the 1341 cm^{-1} ROA band in aldolase is more prominent than in dehydroquinase. Such differences may be due to sample quality, which results in a degradation of the signal to noise ratio. As with aldolase the absence of a negative 1360 cm^{-1} ROA band would suggest that little or no β -turns are present and lends support to the idea that parallel β -strands are present in the core of this biomolecule. In the amide I region of the ROA spectrum a negative/positive couplet centred at \sim

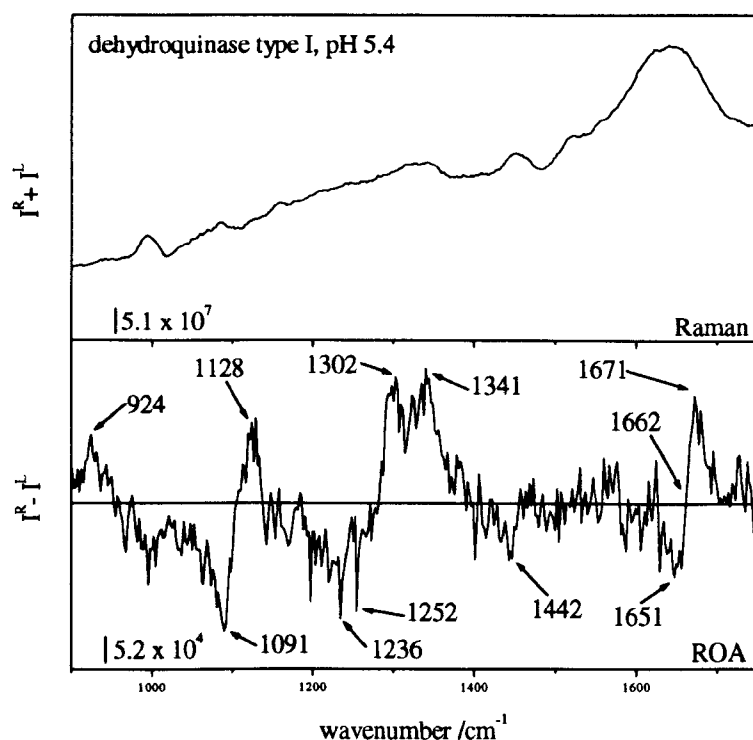


Figure 8.15 Raman and ROA spectra of type I dehydroquinase in H_2O .

1662 cm^{-1} is similarly positioned to that of aldolase, with a negative ROA band at $\sim 1651\text{ cm}^{-1}$ and a positive ROA band at $\sim 1671\text{ cm}^{-1}$. However, TIM barrels have equal amounts of α -helix and β -sheet (see aldolase). Consequently, it may be expected that the positive 1671 cm^{-1} band should be located at a slightly lower wavenumber at ~ 1667 to 1669 cm^{-1} . Clearly, further studies are needed to better demarcate the differences found in proteins with differing amounts of α -helix and β -sheet structures.

8. 2. Ordered and Disordered Polysaccharides

8.2.1. Polysaccharides

The polysaccharide samples of dextran, glycogen and pullulan are all thought to adopt random coil conformations in aqueous solutions (Bell, 1994), yet they reveal that their ROA spectra are richly detailed, unlike some of the random coil proteins

seen in the last chapter. Laminarin, however, is ordered and adopts a triple helix structure which is found in the solid state (Bell *et al.*, 1994; Bell *et al.*, 1995). Note that the spectra shown in this section were obtained in collaboration with Dr. Alasdair F. Bell.

8. 2. 2. Dextran, Glycogen, Pullulan and Laminarin

The secondary structure of dextran consists of a series of $\alpha(1-6)$ linked D-glucose units, with $\sim 5\%$ $\alpha(1-3)$ linkages that form branch points (Sidebotham, 1974), and has a D-isomaltose repeat unit (Bell, 1994). Glycogen consists mostly of $\alpha(1-4)$ linkages (repeat unit: D-maltose (Bell, 1994)), but also has $\alpha(1-6)$ linkages occurring at branch points every 10 to 14 residues (Greenwood, 1970). Pullulan, however, is completely linear consisting of D-maltotriose repeat units linked by $\alpha(1-6)$ linkages (Bell, 1994; Nishinari *et al.*, 1991). The disordered nature of these polysaccharides is thought to originate from flexible main chain $\alpha(1-6)$ linkages in dextran and pullulan, and from frequent branching in glycogen (Bell, 1994). Laminarin is a $\beta(1-3)$ linked species with a number average degree of polymerisation of ~ 26 to 31 with an average chain length of 7 to 10 units and has two or three branch points with $\beta(1-6)$ linkages per chain (Bell, 1994; Percival & McDowell, 1985; Bull & Chesters, 1996; Nelson & Lewis, 1974).

The Raman and ROA spectra of dextran and pullulan are presented in Figure 8.16, with glycogen and laminarin presented in Figure 8.17. In the ROA spectra of polysaccharides the glycosidic linkage is the easiest signal to assign, as polysaccharides have approximately one glycosidic linkage per residue, whereas, for example, a dimer has one glycosidic linkage for every two residues (Bell, 1994). Therefore, polysaccharide samples have a higher proportion of glycosidic linkages which become more prominent in their ROA spectra (Bell, 1994). In the ROA spectrum of dextran a negative band at $\sim 896\text{ cm}^{-1}$ and a positive band at $\sim 955\text{ cm}^{-1}$ may represent the $\alpha(1-6)$ glycosidic linkage, as this signal is absent from the ROA spectrum of the repeat unit D-isomaltose (Bell, 1994). For glycogen and pullulan this couplet is centred at $\sim 915\text{ cm}^{-1}$ but is reversed in sign for both of these samples

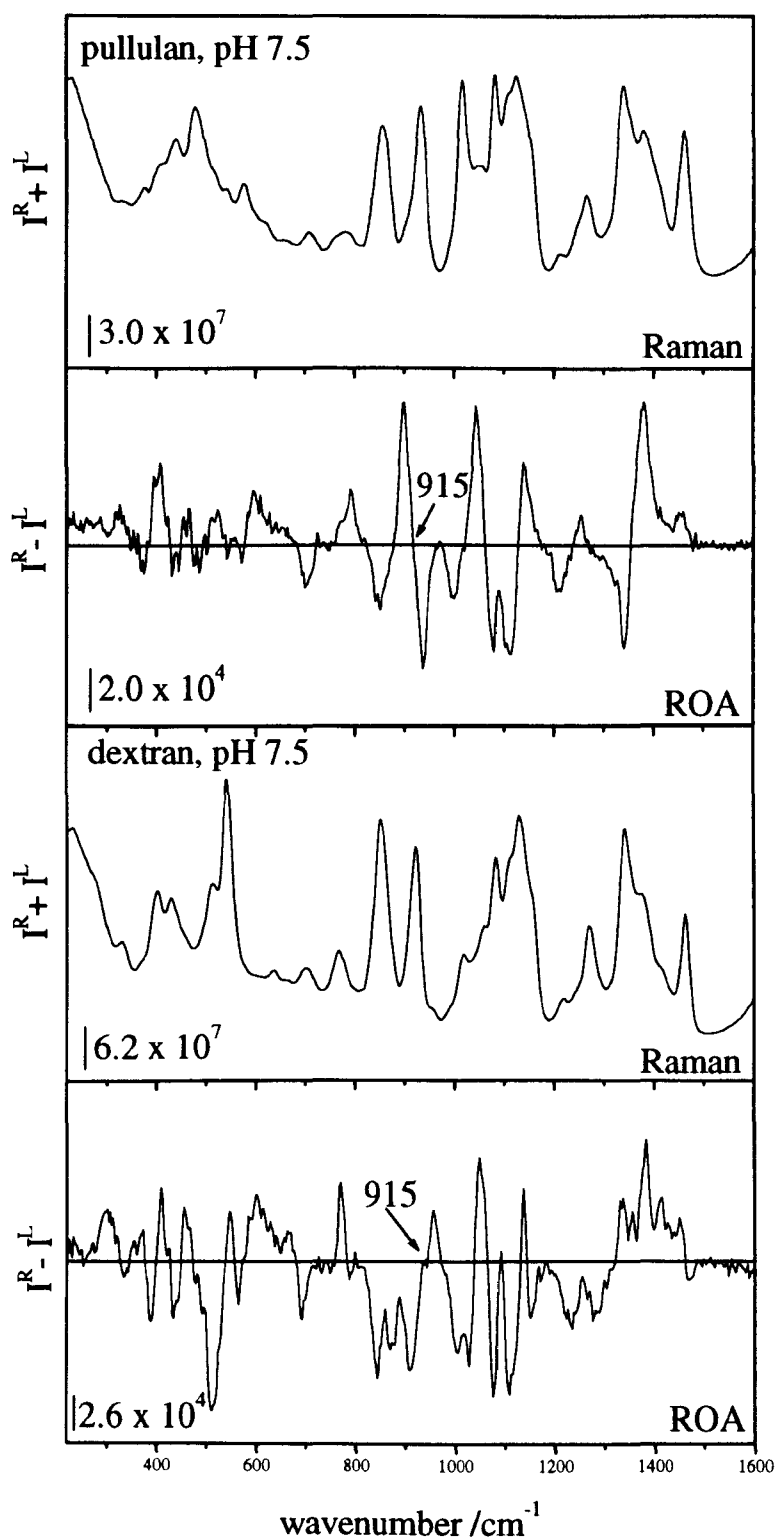


Figure 8.16 Raman and ROA spectra of pullulan (top) and dextran (bottom).

to give a positive-negative couplet and is assigned as originating in the glycosidic $\alpha(1-4)$ linkage of D-maltose, the repeat unit of glycogen (Bell, 1994), whereas pullulan has a $\alpha(1-6)$ linkage (Bell, 1994) from the D-maltotriose repeat unit. In the ROA spectrum of laminarin there are two regions of interest at $\sim 1000\text{ cm}^{-1}$ to 1150 cm^{-1} and at $\sim 1200\text{ cm}^{-1}$ to 1500 cm^{-1} . In the 1000 cm^{-1} to 1150 cm^{-1} region of the Raman and IR spectrum of disaccharides, C-O and C-C stretching vibrations associated with C-O glycosidic oxygen, have been assigned by normal coordinate analysis (Deuchez *et al.*, 1994). When laminarin is compared to its repeat unit, laminaribiose, four main differences are seen in their ROA spectra, with a positive band at $\sim 1033\text{ cm}^{-1}$, a negative band at $\sim 1060\text{ cm}^{-1}$, a positive band at $\sim 1111\text{ cm}^{-1}$ and a negative band at $\sim 1128\text{ cm}^{-1}$ (Bell, 1994). In the ROA spectrum of laminaribiose each of these bands are also present but have the opposite sign pattern, which has been ascribed to the triple helix formation of the laminarin sample (Bell, 1994; Bell *et al.*, 1995). In the ROA spectrum of laminarin between $\sim 1200\text{ cm}^{-1}$ and 1500 cm^{-1} a negative peak at $\sim 1430\text{ cm}^{-1}$ appears and is thought to originate in angle bending deformations about carbon atoms in the glycosidic linkage (Bell, 1994; Bell *et al.*, 1995). The ROA spectrum of laminaribiose has a much weaker negative band at $\sim 1421\text{ cm}^{-1}$ (Bell *et al.*, 1994; Bell *et al.*, 1995). The increased intensity of this band in laminarin is used as an indicator of the glycosidic nature of this band assignment (Bell, 1994).

8.3. Discussion

Due to the difficulty in obtaining naturally occurring proteins with both a glycosylated and non-glycosylated polypeptide backbone, it is difficult to draw many conclusions on the effects of glycosylation with respect to protein secondary structure formation. Deglycosylation studies were considered, to monitor the effects of glycosylation on the polypeptide structure, but were ruled out due to the high cost associated with this process (see chapter 4). Studies on the disordered and ordered polysaccharides of dextran, pullulan, glycogen and laminarin show that although prominent features are seen in both structural types, high sample concentrations are required (~ 50 to 200 mg/ml) to obtain these spectra, whereas, in glycoproteins, the

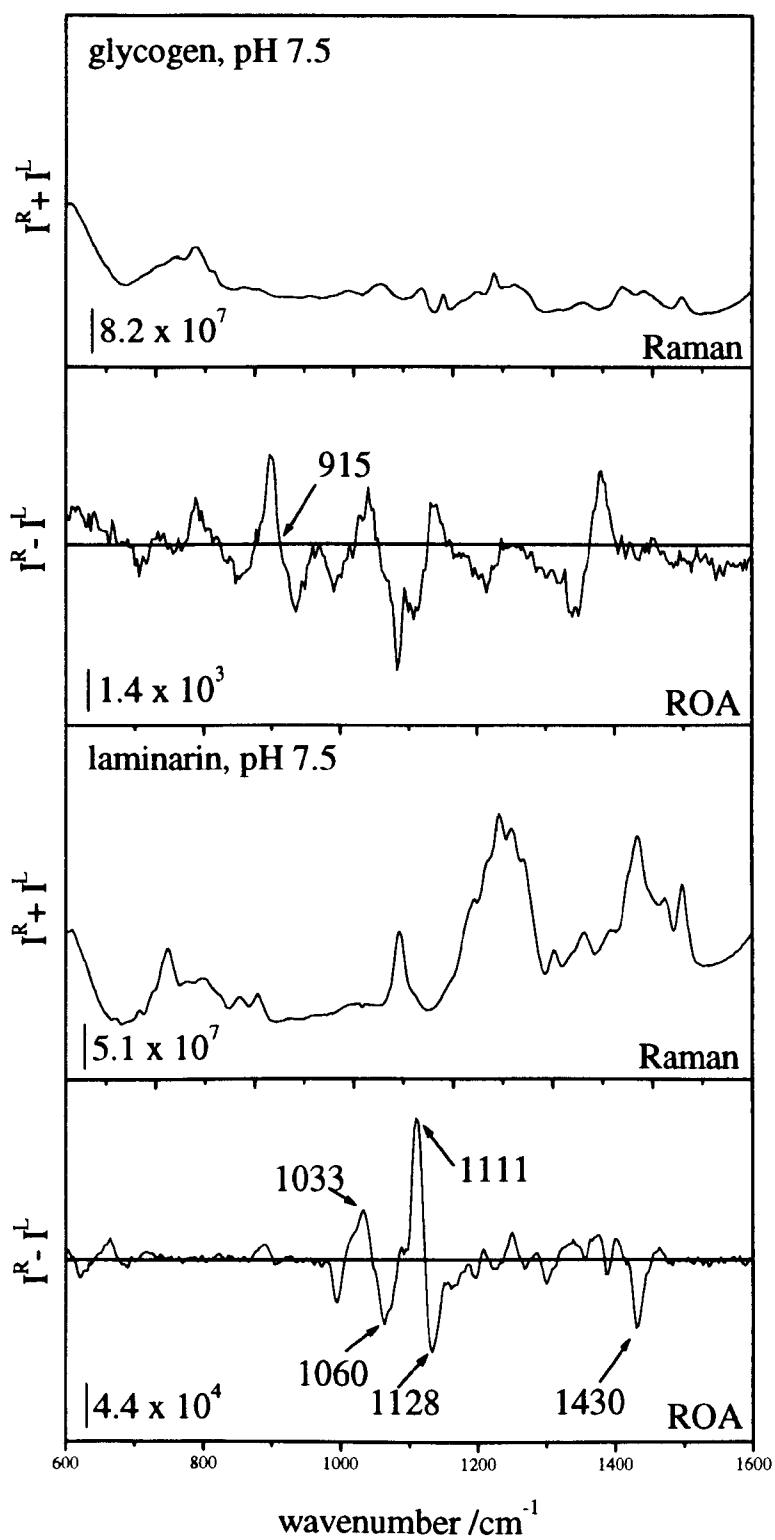


Figure 8.17 Raman and ROA spectra of glycogen (top) and laminarin (bottom).

concentration of glycans is normally much lower (~ 10 to 20% of the total mass, i.e. if the glycoprotein concentration is ~ 70 mg/ml, then the carbohydrate concentration is ~ 7 to 14 mg/ml). Therefore, it is not surprising to find that carbohydrate bands are difficult to detect. A similar problem also exists for conventional Raman spectroscopy (Prescott *et al.*, 1986). The ROA spectra of aldolase and concanavalin A are potentially useful as reference spectra for future studies involving the use of techniques such as regression analysis, where differences in ROA spectra may be used to assign structural motifs to protein samples that may be difficult to study by the conventional techniques of X-ray crystallography and NMR. Aldolase has also proved useful in assigning the negative 1245 cm^{-1} ROA band as originating in β -strand, whereas, in the past, this band was assigned as originating in PPII loop structure (Wilson, 1996; Barron *et al.*, in press). In addition, the spectra of aldolase, HSA and BSA are useful in assigning the positive 1340 cm^{-1} ROA band as hydrated α -helix (Barron *et al.*, in press), whereas formerly it was thought to be 3_{10} -helix (Wilson, 1996; Barron *et al.*, in press). Also, it is informative to compare the ROA spectra of HSA and glycosylated (?) α_1 -antiproteinase, as these spectra show that caution is needed in ensuring sample purity when running spectra. When comparing the Raman and ROA spectra of insulin, the region at ~ 1280 cm^{-1} to 1320 cm^{-1} is ambiguous as Krimm has suggested (Krimm & Bandekar, 1980, 1986; Krimm, 1987) and is dominated by both α -helix vibrations as well as β -turn vibrations, meaning that care is needed when making assignments in this region of both the Raman and ROA spectrum. Finally, a future theme of ROA studies will concentrate on making quantitative comparisons between proteins such as aldolase and dehydroquinase, which have the same fold architecture. Due to the sample quality of dehydroquinase, it was not possible to make a definitive comparison between these samples. However, the ROA spectra of aldolase and dehydroquinase would appear to be more similar than dissimilar, at least in a qualitative sense.

References

- Abraham, C. R. & Potter, H., in *Serine Protease and Their Serpin Inhibitors in The Nervous System*, ed. Festoff, B. W., Plenum Press, New York, 1990, 321.
- Adachi, N., in *Serine Protease and Their Serpin Inhibitors in The Nervous System*, ed. Festoff, B. W., Plenum Press, New York, 1990, 285.
- Adad-Zapatero, C., Rydel, T. J. & Erickson, J., *Proteins* **8** (1990) 62.
- Adzhubei, A. A., Eisenmenger, F., Tumanyan, V. G., Zinke, M., Brodzinski, S. & Esipova, *Biochem. Biophys. Res. Commun.* **146** (1987) 934.
- Akivis, M. A. & Goldberg, V. V., *An Introduction to Linear Algebra and Tensors*, Dover Publications, Inc., New York, 1977.
- Amos, R. D. & Rice, J. E., 'CADPAC', *The Cambridge Analytical Derivatives Package*, Cambridge, Issue 4.0, 1987.
- Amos, R. D., *Chem. Phys. Letts.* **87** (1982) 24.
- Anfinsen, C. B., *Science* **181** (1973) 223.
- Arnold, U. & Ulbrich-Hormann, R., *Biochem.* **36** (1997) 2166.
- Arseniev, A., Schultze, P., Wörgötter, E., Braun, W., Wagner, G., Vašák, M., Kägi, J. H. R. & Wüthrich, K., *J. Mol. Biol.* **201** (1988) 637.
- Artymiuk, P. J., Grindley, H. M., Poirrette, A. R., Rice, D. W., Ujah, E. C. & Willett, P., *J. Chem. Information and Computer Sci.* **34** (1994) 54.
- Asher, S. A., Chi, Z. H. & Li, P. S., *J. Raman Spectrosc.* **29** (1998) 927.
- Athès., V., Combes, D. & Zwick, A., *J. Raman Spectrosc.* **29** (1998) 373.
- Atkins, P. W. & Barron, L. D., *Molec. Phys.* **16** (1969) 453.
- Atkins, P. W. & Friedman, R. S., *Molecular Quantum Mechanics*, 3rd Edition, Oxford University Press, Oxford, 1997.
- Badger, J., Harris, M. R., Reynolds, C. D., Evans, A. C., Dodson, E. J., Dodson, G. G. & North, A. C. T., *ACTA Crystallogr. (Section B)* **47** (1991) 127.
- Barron, L. D. & Buckingham A. D., *Mol. Phys.* **20** (1971) 1111.
- Barron, L. D. & Buckingham, A. D., *Ann. Rev. Phys. Chem.* **26** (1975) 381.

- Barron, L. D. & Buckingham, A. D., *J. Am. Chem. Soc.* **96** (1974) 4769.
- Barron, L. D. & Hecht, L., in *Circular Dichroism: Principles and Applications*, eds. Nakanishi, K., Berova, N. & Woody, R. W., VCH Publishers, New York, 1994, 179.
- Barron, L. D. & Vrbancich, J., *J. Raman Spectrosc.* **15** (1984) 47.
- Barron, L. D., Bogaard, M. P. & Buckingham, A. D., *J. Am. Chem. Soc.* **95** (1973) 603.
- Barron, L. D., Gargaro, A. R. & Wen, Z. Q., *J. Chem. Soc.* (1990a) 1034.
- Barron, L. D., Gargaro, A. R., Hecht, L. & Polavarapu, P. L., *Spectrochim. Acta* **47A** (1991) 1001.
- Barron, L. D., Gargaro, A. R., Hecht, L., Polavarapu, P. L. & Sengeta, H., *Spectrochim. Acta* **48A** (1992) 1051.
- Barron, L. D., Hecht, L., Bell, A. F. & Wilson, G., *Appl. Spectrosc.* **50** (1996) 619.
- Barron, L. D., Hecht, L., Blanch, E. W. & Bell, A. F., *Prog. Biophys. Mol. Biol.*, in press.
- Barron, L. D., Hecht, L., Gargaro, A. R. & Hug, W., *J. Raman Spectrosc.* **21** (1990b) 375.
- Barron, L. D., *Molecular Light Scattering and Optical Activity*, Cambridge University Press, Cambridge, 1982.
- Barskaya, T. V. & Ptitsin, O. B., *Biopolymers* **10** (1971) 2181.
- Becker, J. W., Reeke, G. N., Wang, J. L., Cunningham, B. A. & Edelman, G. M., *J. Biol. Chem.* **250** (1975) 1513.
- Belasco, J. G. & Knowles, J. R., *Biochem.* **19** (1980) 472.
- Bell, A. F., *Doctoral Thesis*, University of Glasgow, 1994.
- Bell, A. F., Hecht, L. & Barron, L. D., *J. Am. Chem. Soc.* **116** (1994) 5155.
- Bell, A. F., Hecht, L. & Barron, L. D., *J. Raman Spectrosc.* **24** (1993) 633.
- Bell, A. F., Hecht, L. & Barron, L. D., *J. Raman Spectrosc.* **26** (1995) 1071.
- Bentley, R., *Crit. Rev. Biochem. Mol. Biol.* **25** (1990) 110.
- Bergethon, P. R., *The Physical Basis of Biochemistry: The Foundations of Molecular Physics*, Springer, New York, 1998.
- Bernstein, F. C., *et al.*, *J. Mol. Biol.*, **112** (1977) 535.
- Beynon, J. D. E. & Lamb, D. R., *Charge-Coupled Devices and Their Applications*, McGraw-Hill, London, 1980.
- Bi, R. -C., Dauter, Z., Dodson, E., Dodson, G., Giordano, F. Reynolds, C., *Biopolymers*

23 (1984) 391.

Billings, B. H., *J. Opt. Soc. Am.* **41** (1951) 966.

Birke, S. S., Ismailu, A. & Diem, M., *Biochem.* **31** (1992) 450.

Blanch, E. W., Hecht, L. Barron, L. D., *Protein Science* **8** (1999) 1362.

Blom, N. S. & Sygusch, J., *Nat. Struct. Biol.* **4** (1997) 36.

Bode, W. & Epp, O., *Eur. J. Biochem.* **147** (1985) 387.

Bose, P. K., Polavarapu, P. L., Barron, L. D. & Hecht, L., *J. Phys. Chem.* **94** (1990) 1734.

Brandon, C. & Tooze, J., *Introduction to Protein Structure*, Garland Publishing, New York, 1991.

Brange, J., Ribel, U., Hansen, J. F., Dodson, G., Hansen, M. T., Havelund, S., Melberg, S. G., Norris, F., Norris, K., Snel, L., Sorensen, A. R. & Voigt, H. O., *Nature* **333** (1988) 679.

Braun, W., Vašák, M., Robbins, A. H., Stout, C. D., Wagner, G., Kägi, J. H. R. & Wüthrich, K., *Proc. Acad. Sci., U.S.A.* **89** (1992) 10124.

Bruch, M., Weiss, V. Engel, J., *J. Biol. Chem.* **236** (1988) 16626.

Bull, A. T. & Chesters, C. G. C., *Adv. in Enzymology* **28** (1996) 325.

Burley, R. W. & Vadhera, D. V., *The Avian Egg: Chemistry and Biology*, J. Wiley, New York, 1989.

Byrne, B. M., van het Schip, A. D., van de Klunder, J. A., Arnberg, A. C., Gruber, M. & Geert, A. B., *Biochem.* **23** (1984) 4275.

Carey, P. R., *Biochemical Applications of Raman and Resonance Raman Spectroscopy*, Academic Press, New York, 1982.

Carter, C. W., in *Bioorganic Chemistry: Peptides and Proteins*, ed. Hecht, S. M., Oxford University Press, New York, 1998, 153.

Carter, D. C., He, X. -M., Munson, S. H., Twigg, P. D., Gernert, K. M., Broom, M. B. & Miller, T. Y., *Science* **244** (1989) 1195.

Chan, H. S. & Dill, K. A., *Proc. Natl. Acad. Sci., U.S.A.* **87** (1990) 6388.

Chang, R., *Basic Principles of Spectroscopy*, McGraw-Hill, New York, 1971.

Chen, X. G., Schweitzerstenner, R., Asher, S. A., Mirkin, N. G. & Krimm, S., *J. Phys. Chem.* **99** (1995) 3047.

Chothia, C., *Nature* **248** (1974) 338.

- Chothia, C. & Lesk, A. M., *J. Mol. Biol.* **196** (1987) 901.
- Chou, K. -C. & Carlacci, L., *Proteins* **9** (1991) 280.
- Chou, K. -C., *Proteins* **21** (1995) 319.
- Chou, P. Y. & Fasman, G. D., *Biophys. J.* **26** (1979) 367.
- Clarke, D. & Gringer, J. F., *Polarized Light and Optical Measurement*, 1st Edition, Pergamon Press, Oxford, 1971.
- Creamer, T. P., *Proteins* **33** (1998) 218.
- Creighton, T. E., *Proteins*, W. H. Freeman, New York, 1994.
- Creighton, T. E., *Proteins: Structures and Molecular Properties*, 2nd Edition, W. H. Freeman, New York, 1993.
- Darby, N. J. & Creighton, T. E., in *Protein Structure*, ed. Rickwood, D., IRL Press, Oxford University Press, Oxford, 1993.
- David, S., *The Molecular and Supramolecular Chemistry of Carbohydrates: A Chemical Introduction to the Glycosciences*, Oxford University Press, Oxford, 1997.
- Davidson, B. & Fasman, G. D., *Biochem.* **6** (1967) 1616.
- de la Sierra, I. L., Quillien, L., Flecker, P., Gueguen, J. Brunie, S., *J. Mol. Biol.* **285** (1999) 1195.
- Deuchez, , M., Derreumaux, P., Sekkal, M., Lagant, P., Legrand, P. & Vergoten, G., *Spectrochim. Acta* **50A** (1994) 87.
- Devlin, F. J. & Stephens, P. J., *J. Phys. Chem. A* **103** (1999) 527.
- Diem, M., *Introduction to Modern Vibrational Spectroscopy*, J. Wiley, New York, 1993.
- Dill, K. & Chan, H. S. *Nature Struct. Biol.* **4** (1997) 10.
- Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D. & Chan, H. S., *Prot. Sci.* **4** (1995) 561.
- Edmund, A. B., Ely, K. R., Sly, D. A., Westholm, F. A., Powers, D. A. & Liener, I. E., *Biochem.* **10** (1971) 3554.
- Elliott, P. R., Abrahams, J. & Lomas, D. A., *J. Mol. Biol.* **275** (1998) 419.
- Epperson, P. M. & Denton, M. B., *Anal. Chem.* **61** (1990) 1513.
- Escribano, J. R., *Chem. Phys. Lett.* **121** (1985) 191.
- Evans, J. N. S., *Biomolecular NMR Spectroscopy*, Oxford University Press, Oxford, 1995.

- Fagnano, C. & Fini, G., *J. Mol. Struct.* **294** (1993) 111.
- Falkin, D. & Vosloo, M., *Spectroscopy Europe* **5** (1993) 16.
- Flöckner, H., Braxenthaler, M., Lackner, P., Jaritz, M., Ortner, M. & Sippl, M. J., *Proteins* **23** (1995) 376.
- Frautschy, S. A., Horn, D. L., Sigel, J. J., Harris-White, M. E., Mendoza, J. J., Yang, F., Saïdo, T. C. Cole, G. M., *The Journal of Neuroscience* **18** (1998) 8311.
- Frushour, B. G. & Koenig, J. L., *Biopolymers* **13** (1974) 1809.
- Fujinaga, M., Sielecki, A. R., Read, R. J., Ardelt, W., Laskowski, M. & James, M. N. G., *J. Mol. Biol.* **195** (1987) 397.
- Fukuda, M., *Molecular Glycobiology*, eds. Fukuda, M. & Hindsgaul, O., IRL Press Oxford University, Oxford, 1994, 1.
- Gargaro, A. R., Barron, L. D. & Hecht, L., *J. Raman Spectrosc.* **24** (1993) 91.
- Gast, K., Damaschun, H., Eckert, K., Schulze-Forester, K., Maurer, R., Müller-Frohne, M., Zirwer, D., Czarnecki, J. & Damaschun, G., *Biochem.* **34** (1995) 13211.
- Gerrard, A. & Burch, J. M., *Introduction to Matrix Methods in Optics*, Dover, New York, 1994.
- Gourley, D. G., Shrive, A. K., Polikarpov, I., Krell, T., Coggins, J. R., Hawkins, A. R., Isaacs, N. W. & Sawyer, L., *Nat. Struct. Biol.* **6** (1999) 1072.
- Green, N. M., *Adv. Protein Chem.* **29** (1975) 85.
- Greenfield, N. & Fasman, G. D., *Biochem.* **8** (1969) 4108.
- Greenwood, C. T., *The Carbohydrates*, eds. Pigman, W. & Horton, D., Academic Press, London, 1970.
- Gudehus, D. H. & Hegyi, D. J., *The Astronomical Journal* **90** (1985) 130.
- Harris, J. M., Kleanthous, C., Coggins, J. R., Hawkins, A. R., Abell, C., *J. Chem. Soc. Chem. Commun.* (1993) 1080.
- Harris, J. M., Watkins, W. J., Hawkins, A. R., Coggins, J. R. & Abell, C., *J. Chem. Soc., Perkin Trans. 1* (1996) 2371.
- He, X. -M. & Carter, D. C., *Nature* **358** (1992) 209.
- Hecht, L. & Barron L. D., *Appl. Spectrosc.* **44** (1990a) 483.
- Hecht, L. & Barron, L. D., *Appl. Spectrosc.* **72** (1990b) 441.
- Hecht, L. & Nafie, L. A., *Chem. Phys. Letts.* **174** (1990) 575

- Hecht, L., & Barron, L. D., *Faraday Discuss. Chem. Soc.* **99** (1994) 35.
- Hecht, L., Barron, L. D., Blanch, E. W., Bell, A. F. & Day, L. A., *J. Raman Spectrosc.* **30** (1999) 815.
- Hecht, L., Barron, L. D., Gargaro, A. R., Wen, Z. Q. & Hug, W. J., *J. Raman Spectrosc.* **23** (1992a) 401.
- Hecht, L., Barron, L. D., *J. Mol. Struct.* **347** (1995) 449.
- Hecht, L., Che, D. & Nafie, L. A., *Appl. Spectrosc.* **45** (1991) 18.
- Hecht, L., Che, D. & Nafie, L. A., *J. Phys. Chem.* **96** (1992b) 4266.
- Hecht, L., Jordanov, B. & Schrader, B., *Appl. Spectrosc.* **41** (1987) 295.
- Herald, T. J. & Smith, D. M., *J. Agric. Food Chem.* **40** (1992) 1737.
- Holzwarth, G. & Doty P., *J. Am. Chem. Soc.* **87** (1965) 218.
- Holzwarth, G., Hsu, E. C., Mosher, H. S., Faulkner, T. R. & Moscowitz, A., *J. Am. Chem. Soc.* **96** (1974) 251.
- Hoogstraten, C. G., Choe, S., Westler, W. M. & Markley, J. L., *Pro. Sci.* **4** (1995) 2289.
- Honzatko, R. B. & Williams, R. W., *Biochem.* **21** (1982) 6201.
- Hsu, I. N., Delbaere, L. T., James, M. N. G. & Hofmann, T., *Nature* **266** (1977) 140.
- Huber, R. & Carrell, R. W., *Biochem.* **28** (1989) 8951.
- Hug, W., *Appl. Spectrosc.* **35** (1981) 115.
- Hutchinson, E. G. & Thornton, J. M., *Prot. Sci.* **3** (1994) 2207.
- Janin, J., Wodak, S. J., Levitt, M. & Maigret, B. *J. Mol. Biol.* **125** (1978) 357.
- Jefferey, G. A. & Saenger, W., *Hydrogen Bonding in Biological structures*, Springer-Verlag, Berlin, 1994.
- Joao, H. C. & Dwek, R. A., *Eur. J. Biochem.* **218** (1993) 239.
- Jørgensen, A. M. M., Kristensen, S. M., Led, J. J. & Balschmidt, P., *J. Mol. Biol.* **261** (1996) 341.
- Kabsch, W. & Sander, C., *Biopolymers* **22** (1983) 2577.
- Kaiser Optical Systems, Inc., *Specification Manual*, USA.
- Karpen, M. E., de Haseth, P. L. & Neet, K. E., *Prot. Sci.* **1** (1992) 1333.
- Keiderling, T. A., in *Circular Dichroism and the Conformational Analysis of Biomolecules*, ed. Fasman, G., Plenum Press, New York, 1996, 555.
- Keiderling, T. A., in *Circular Dichroism: Principles and Applications*, eds. Nakanishi,

- K., Berova, N. & Woody, R. W., VCH Publishing, New York, 1994, 497.
- Kirkwood, J. G., *J. Chem. Phys.* **5** (1937) 479.
- Kleanthous, C., Deka, R., Davis, S. M., Kelly, A., Cooper, A., Harding, S. E., Price, N. C., Hawkins, A. R., & Coggins, J. R., *Biochem. J.* **282** (1992) 687.
- Kliger, D. S., Lewis, J. S. & Randall, C. E. *Polarized Light in Optics and Spectroscopy*, Academic Press, London, 1990.
- Kobayashi, N., Yamato, T. Go, N., *Proteins* **28** (1997) 109.
- Kraulis, P. J., *J. Appl. Cryst.* **24** (1991) 946.
- Krimm, S. & Bandekar, J., *Adv. Protein Chem.* **38** (1986) 181.
- Krimm, S. & Bandekar, J., *Biopolymers* **19** (1980) 1.
- Krimm, S. & Reisdorf, W. C., *Faraday Discuss.* **99** (1994) 181.
- Krimm, S., in *Biological Applications of Raman Spectroscopy Vol. 1*, ed. Spiro, T. G., J. Wiley, New York, 1987, 1.
- Kurokawa, H., Mikami, B & Hirose, M., *J. Mol. Biol.* **254** (1995) 196.
- Kusumoto, Y., Lomakin, A., Peplow, D. B. Benedek, G. B., *Proc. Natl. Acad. Sci., U.S.A.* **95** (1998) 12277.
- Laskowski, M & Kato, I., *Ann. Rev. Biochem.* **49** (1980) 593.
- Lasso, J. N., Bogumil, R. & Nakai, S., *Comp. Biochem. Physiol.* **106B** (1993) 919.
- Lesk, A. M., *Protein Architecture*, IRL Press, Oxford University Press, Oxford, 1991.
- Levitt, M. & Chothia, C., *Nature* **261** (1976) 552.
- Lightner, D., in *Circular Dichroism: Principles and Applications*, eds. Nakanishi, K., Berova, N. & Woody, R. W., VCH Publishers, New York, 1994, 259.
- Lillo, M. P., Szpikowska, B. K., Mas, M. T. & Beechem, J. M., *Biochem.* **36** (1997) 11273.
- Linderstøm-Lang, K. U., *The Lane Medical Lectures*, Stanford University Press, Stanford, CA, 1952.
- Lippert, J. L., Tymininski, D. & Desmeules, P. J., *J. Am. Chem. Soc.* **98** (1976) 7075.
- Livnah, O., Bayer, E. A., Wilchek, M Sussman, J. L., *Proc. Natl. Acad. Sci., U.S.A* **90** (1993) 5076.
- Lobermann, H., Torkuoka, R., Deisenhofer, J & Huber, R., *J. Mol. Biol.* **117** (1984) 531.
- Loeber, A. P., *J. Opt. Soc. Am.* **72** (1982) 650.

- Lomas, D. A., Elliott, P, Chang, W. W., Wardell, M. R. & Carrell, W., *J. Biol. Chem.* **270** (1995) 5282.
- Lord, R. C. Yu, N. –T., *J. Mol. Biol.* **50** (1970) 509.
- Marquart, M. & Deisenhofer, J., *Immunology Today* **3** (1982) 160.
- Matsuda, T., Watanabe, K. & Sato, Y., *F.E.B.S. Letters* **124** (1981) 185.
- Mellet, P., Michels, B. & Bieth, J. G., *J. Biol. Chem.* **271** (1996) 30311.
- Messerle, B. A., Schäffer, A., Vašák, M, Kägi, J. H. R. & Wüthrich, K., *J. Mol. Biol.* **214** (1990) 765.
- Miyazawa, T., Shimanouchi, T. & Mizushima, S., *J. Chem. Phys.* **29** (1958) 611.
- Mochizuki, K., *Appl. Opt.* **23** (1984) 3284.
- Moffitt, W. & Yang, J. T., *Proc. Nat. Acad. Sci., U.S.A.* **42** (1956) 596.
- Muñoz, V. & Serrano, L., *Proteins* **20** (1994) 301.
- Nafie, L. A., *4th Int. Conference on C. D.*, Bochum, FRG, 1991, 101.
- Nafie, L. A. & Freeman, T. B., *Chem. Phys. Letts.* **154** (1989) 260
- Nafie, L. A. & Freeman, T. B., *J. Phys. Chem.* **78** (1983) 7180.
- Nafie, L. A. & Zimba, C. G., in *Biological Applications of Raman Spectroscopy, Vol. 1*, ed. Spiro, T. G., J. Wiley, New York, 1987, 307.
- Nafie, L. A., *Appl. Spectrosc.* **50** (1996) 14A.
- Nafie, L. A., Che, D., Yu, G. S. & Freeman, T. B., in *Biomolecular Spectroscopy II, Proc. SPIE 1432*, eds. Akhmanov, S. A., Poroshina, N. I. Koroteev, N. I. & Toluetaev, B., 1991, 441.
- Nafie, L. A., Cheng, J. C. & Stephens, P. J., *J. Am. Chem. Soc.* **97** (1975) 3842.
- Nelson, T. E. & Lewis, B. A., *Carbohydr. Res.* **33** (1974) 63.
- Nishinari, K., Kohyama, K., Williams, P. A., Phillips, G. O., Burchard, W. & Ogino, K., *Macromolecules*, **24** (1991) 5590.
- Olszewski, K. A., Kolinski, A. & Skolnick, J., *Protein Engineering* **9** (1996) 5.
- Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B. & Thornton, J. M., *Structure* **5** (1997) 1093.
- Overman, S. A. & Thomas, G., *J. Raman Spectrosc.* **29** (1998) 23.
- Painter, P. C. & Koenig, J. L., *Biopolymers* **14** (1975) 457.
- Pande, J., Pande, C., Gilg, D., Vašák, M, Callander, R. & Kägi, J. H. R., *Biochem.* **25**

- (1986) 5526.
- Paterlini, M. G., Freeman, T. B. & Nafie, L. A., *Biopolymers* **25** (1986) 1751.
- Pauling, L. & Corey, R. B., *Proc. Nat. Acad. Sci., U.S.A.* **37** (1951a) 235.
- Pauling, L. & Corey, R. B., *Proc. Nat. Acad. Sci., U.S.A.* **37** (1951b) 251.
- Pemberton, J. E., Sobocinski, R. L., Bryant, M. A. & Carter, D. A., *Spectroscopy International* **2** (1990) 26.
- Percival, E. & McDowell, R. H., *Biochemistry of Storage Carbohydrates in Green Plants*, eds. Dey, P. M. & Dixon, R. A., Academic Press, London, 1985.
- Perczel, A. Jákli, I., *Biopolymers* **38** (1995) 723.
- Perutz, M. F., *Mechanism of Cooperativity and Allosteric Regulation in Proteins*, Cambridge University Press, Cambridge, 1990.
- Perutz, M., *Protein Structure*, W. H. Freeman, New York, 1992.
- Pézolet, M., Pigeon-Gosselin, M. & Coulombe, L., *Biochemica et Biophysica Acta* **453** (1976) 502.
- Polavarapu, P. L. & Deng, Z., *Appl. Spectrosc.* **50** (1996) 686.
- Polavarapu, P. L., *J. Phys. Chem.* **94** (1990) 8106.
- Polavarapu, P. L., *Vibrational Spectra: Principles and Applications with Emphasis on Optical Activity*, Elsevier, Amsterdam, 1998.
- Polavarapu, P.L. Black, T. M., Barron, L. D. & Hecht, L., *J. Am. Chem. Soc.* **115** (1993a) 7736.
- Polavarapu, P.L. Bose, P.K., Hecht, L. & Barron, L. D., *J. Chem. Phys.* **97** (1993b) 11211.
- Pons, T., Olema, O., Chinea, G., Beldarraín, A., Márquez, G., Acosta, N., Rodríguez, L. & Valencia, A., *Proteins* **33** (1998) 383.
- Prescott, B., Renugopalakrishnan, V., Glimcher, M. J., Bhushan, A. & Thomas, G., *Biochem.* **25** (1986) 2792.
- Privalov, P. L., in *Protein Folding*, ed. Creighton, T. E., W. H. Freeman, New York, 1992, 83.
- Pugliese, L., Coda, A., Malcovati, M. & Bolognesi, M., *J Mol. Biol.* **178** (1984) 787.
- Rademacher, T. W., Jaques, A. & Williams, P. J., in *Abnormalities of IgG Glycosylation and Immunological Disorders*, eds. Isenberg, D. A. & Rademacher, T. W., Wiley,

- New York, 1996, 1.
- Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V., *J. Mol. Biol.* **7** (1963) 95.
- Reeke, G. N., Becker, J. W. Edelman, G. M., *J. Mol. Biol.* **250** (1975) 1525.
- Richards, F. M. & Lim, W. A., *Quart. Rev. Biophys.* **26** (1993) 423.
- Richardson, J. S., *Nature* **268** (1977) 495.
- Roa, J. S., Suzuki, R. & Festoff, B. W., in *Serine Protease and Their Serpin Inhibitors in The Nervous System*, ed. Festoff, B. W., Plenum Press, New York, 1990, 301.
- Robbins, A. H., McRee, D. E., Williamson, M., Collett, S. A., Xuong, N. H., Furey, W. F., Wang, B. C. & Stout, C. D., *J. Mol. Biol.* **221** (1991) 1269.
- Rogers, A. & Nordén, B., *Circular Dichroism and Linear Dichroism*, Oxford Chemistry Masters, Oxford University Press, Oxford, 1997.
- Roitt, I., *Roitt's Essential Immunology*, 9th Edition, Blackwell Science Ltd., Oxford, 1997.
- Rose, G. D., Gierasch, L. M. & Smith, J. A., in *Adv. Prot. Chem.* **37** (1985) 1.
- Rudd, P. M. & Dwek, R. A., *Critical Reviews in Biochemistry and Molecular Biology*, **32** (1997) 1.
- Rudd, P. M., Joao, H. C., Coghill, E., Fiten, P., Saunders, M. R., Opdenakker, G. & Dwek, R. A., *Biochem.* **33** (1994) 17.
- Saul, F. A. & Poljack, R. J., *Proteins* **14** (1991) 363.
- Sawaya, R., in *Serine Protease and Their Serpin Inhibitors in The Nervous System*, ed. Festoff, B. W., Plenum Press, New York, 1990, 293.
- Schneider, D. J., Roe, A. L., Mayer, R. J. & Que, L., *J. Biol. Chem.* **259** (1984) 9699.
- Schülke, N. & Schmid, F. X., *J. Biol. Chem.* **263** (1988) 8832.
- Scully, J. & Hermans, J., *J. Mol. Biol.* **235** (1994) 682.
- Schultze, P., Wörgötter, E., Braun, W., Wagner, G., Vašák, M., Kägi, J. H. R. & Wüthrich, K., *J. Mol. Biol.* **203** (1988) 251.
- Seikagaku, *AMS Biotechnology (Europe) Ltd.*, 1999 Catalog.
- Shriver, D. F., Atkins, P. W. & Langford, C. H., *Inorganic Chemistry*, 2nd Edition, Oxford University Press, Oxford, 1994.
- Sidebotham, R. L., *Adv. in Carbohydr. Chem. Biochem.* **30** (1974) 371.
- Smith, L. J., Fiebig, K. M., Schwalbe, H. & Dobson, C. D., *Fol. Des.* **1** (1996) R95.

- Smyth, E., Syme, C. D., Blanch, E. W., Hecht, L., Vašák, M. & Barron, L. D., to be published.
- Snatzke, G., in *Circular Dichroism: Principles and Applications*, eds. Nakanishi, K., Berova, N. & Woody, R. W., VCH Publishers, New York, 1994, 1.
- So, L. L. & Goldstein, I. J., *Biochim. Biophys. Acta* **165** (1968) 398.
- Spencer, K. M., Freeman, T. B. & Nafie, L. A., *Chem. Phys. Letts.* **149** (1988) 367.
- Stein, P. E., Leslie, A. G. W., Finch, J. T. & Carrell, R. W., *J. Mol. Biol.* **221** (1991).
- Stone, A. J., *Molec. Phys.* **33** (1976) 293.
- Subramanian, E., Swan, I. D. A., Liu, M., Davies, D. R., Jenkins, J. A., Tickle, I. J. & Blundell, T. L., *Proc. Natl. Acad. Sci., U.S.A.* **74** (1977) 556.
- Surbeck, H., Hug, W., Gremand, M., Bridoux, M., Defontaine, A. & De Silva, E., *Opt. Commun.* **38** (1981) 57.
- Syngusch, J., Beaudry, D. & Allaire, M., *Proc. Natl. Acad. Sci., U.S.A.* **84** (1987) 7846.
- Takada, K., Chida, K. & Noda, J., *J. Opt. Soc. Am.* **A11** (1988) 1905.
- Tanzi, R. E., *Serine Protease and Their Serpin Inhibitors in The Nervous System*, ed. Festoff, B. W., Plenum Press, New York, 1990, 313.
- Tedesco, J. M., Owen, H., Pallister, D. M. & Morris, M.D., *Anal. Chem.* **65** (1993) 411A.
- Teraoka, J., Bell, A. F., Hecht, L. & Barron, L. D., *J. Raman Spectrosc.* **29** (1998) 67.
- Thorne, J. L., Goldman, N. & Jones, D. T., *Mol. Biol. Evol.* **13** (1996) 666.
- Thotakura, N. R. & Bahl, O. P., *Methods in Enzymology* **138** (1987) 350.
- Tiffany, M. L. & Krimm S., *Biopolymers* **8** (1969) 347.
- Tiffany, M. L. & Krimm, S., *Biopolymers* **6** (1968) 1379.
- Toniolo, C., Polese, A., Formaggio, F., Crisma, M. & Kamphuis, J., *J. Am. Chem. Soc.* **118** (1996) 2744.
- Torreggiani, A. & Fini, G., *J. Raman Spectrosc.* **29** (1998) 229.
- Trimble, R. B. & Atkinson, P. H., *J. Biol. Chem.* **261** (1986) 9815.
- Trimble, R. B., Maley, F. & Chu, F. K., *J. Biol. Chem.* **258** (1983) 2562.
- Tu, A. T., in *Spectroscopy of Biological Systems*, eds. Clark, R. J. H. & Hester, R. E., Wiley, New York, 1986, 47.
- Vlasov, A. P., Kravchuk, Z. I. & Martsev, S. P., *Biochem. (Moscow)* **61** (1996) 155.
- Vogel, H. J., *Biochem.* **22** (1983) 668.

- Vosolov, A. & Woody, R.W., in *Circular Dichroism: Principles and Applications*, eds. Nakanishi, K., Berova, N. & Woody, R. W., VCH Publishers, New York, 1994, 59.
- Wang, J. L., Cunningham, B. A. & Edelman, G. M., *Proc. Natl. Acad. Sci., U.S.A.* **68** (1971) 1130.
- Ware, J. H, Wan, X. S., Rubin, H., Schechter, N. M. & Kennedy, A. R., *Arch. Biochem. Biophys.* **344** (1997) 133.
- Watanabe, K., Matsuda, T & Sato, Y., *Biochemica et Biophysica Acta*, **667** (1981) 242.
- Weber, E., Papamokos, E., Bode, W., Huber, R., Kato, I & Laskowski, M., *J. Mol. Biol.* **158** (1982) 515.
- Weisgerber, S. & Helliwell, J. R., *J. Chem. Soc. Faraday Trans.* **89** (1993) 2667.
- Werner, M. H. & Wemmer, D. E., *Biochem.* **31** (1992) 999.
- Williams, D. H. & Fleming, I., *Spectroscopic Methods in Organic Chemistry*, 4th Edition, McGraw-Hill, London, 1989.
- Williams, R. L., Greene, S. M. & McPherson, A., *J. Biol. Chem.* **262** (1987) 16020.
- Wilson, E. B., Decius, J. C. & Cross, P. C., *Molecular Vibrations*, McGraw-Hill, New York, 1955.
- Wilson, G., *Doctoral Thesis*, University of Glasgow, 1996.
- Wilson, G., Hecht, L. & Barron, L. D., *Biochem.* **35** (1996a) 12518.
- Wilson, G., Hecht, L. & Barron, L. D., *J. Chem. Soc. Faraday Trans.* **92** (1996b) 1503.
- Woody, R. W., *Adv. Biophys. Chem.* **2** (1992) 37.
- Woody, R. W., in *Circular Dichroism: Principles and Applications*, eds. Nakanishi, K., Berova, N. & Woody, R. W., VCH Publishers, New York, (1994) 473.
- Wright Instruments, Ltd., *Instruction Manual*, England.
- Wuyuan, L., Apostol, I., Qasim, M. A., Warne, N., Wynn, R., Zhang, W. L., Anderson, S., Chiang, Y. W. & Laskowski, M., *J. Mol. Biol.* **266** (1997) 441.
- Yarin, J., Kalb, A. J. & Levitzki, A., *Biochim. Biophys. Acta* **165** (1968) 303.
- Yasui, S. C., Pancoska, P., Dukor, R. K., Keiderling, T. A., Renugopalakrishnan, V., Glimcher, M. J. & Clark, R. C., *J. Biol. Chem.* **265** (1990) 3780.
- Zwick, A., Lakhdarghazal, F & Tocanne, J. F., *J. Chem. Soc. Faraday Trans.* **85** (1989) 783.